Preliminary Chapter 1

WHAT IS A FUNCTION?

No concept in mathematics, especially in calculus, is more fundamental than the concept of a function. The term was first used in a 1673 letter written by Gottfried Wilhelm Leibniz, the German mathematician and philosopher who invented calculus independently of Isaac Newton. Since then the term has undergone a gradual extension of meaning.

In traditional calculus a function is defined as a relation between two terms called variables because their values vary. Call the terms x and y. If every value of x is associated with exactly one value of y, then y is said to be a function of x. It is customary to use x for what is called the *independent variable,* and y for what is called the *dependent variable* because its value depends on the value of x.

As Thompson explains in Chapter 3, letters at the end of the alphabet are traditionally applied to variables, and letters elsewhere in the alphabet (usually first letters such as a,b,c. . .) are applied to constants. Constants are terms in an equation that have a fixed value. For example, in $y = ax + b$, the variables are x and y, and a and b are constants. If $y = 2x + 7$, the constants are 2 and 7. They remain the same as x and y vary.

A simple instance of a geometrical function is the dependence of a square's area on the length of its side. In this case the function is called a one-to-one function because the dependency goes both ways. A square's side is also a function of its area.

A square's area is the length of its side multiplied by itself. To express the area as a function of the side, let y be the area, x the side, then write $y = x^2$. It is assumed, of course, that x and y are positive.

A slightly more complicated example of a one-to-one function is the relation of a square's side to its diagonal. A square's diagonal is the hypotenuse of an isosceles right triangle. We know from the Pythagorean theorem that the square of the hypotenuse equals the sum of the squares of the other two sides. In this case the sides are equal. To express the diagonal as a function of the square's side, let y be the diagonal, x the side, and write $y = \sqrt{2x^2}$, or more simply $y = x\sqrt{2}$ to express the side as a function of the diagonal, let y be the side, x the diagonal, and write $y = \sqrt{\frac{x^2}{2}}$, or more simply $y = \frac{x}{\sqrt{2}}$.

The most common way to denote a function is to replace y, the dependent variable, by $f(x)$—f being the first letter of "function." Thus $y = f(x) = x^2$ means that y, the dependent variable, is the square of x. Instead of, say, $y = 2x - 7$, we write $y = f(x) = 2x - 7$. This means that y, a function of x, depends on the value of x in the expression $2x - 7$. In this form the expression is called an *explicit* function of x. If the equation has the equivalent form of $2x - y - 7 = 0$, it is called an *implicit* function of x because the explicit form is implied by the equation. It is easily obtained from the equation by rearranging terms. Instead of $f(x)$, other symbols are often used.

If we wish to give numerical values to x and y in the example $y = f(x) = 2x - 7$, we replace x by any value, say 6, and write $y = f(6) = (2 \cdot 6) - 7$, giving the dependent variable y a value of 5.

If the dependent variable is a function of a single independent variable, the function is called a function of one variable. Familiar examples, all one-to-one functions, are:

The circumference or area of a circle in relation to its radius.

The surface or volume of a sphere in relation to its radius.

The log of a number in relation to the number.

Sines, cosines, tangents, and secants are called trigonometric functions. Logs are logarithmic functions. Exponential functions are functions in which x, the independent variable, is an exponent in a equation, such as $y = 2^x$. There are, of course, endless other examples of more complicated one-variable functions which have been given names.

Functions can depend on more than one variable. Again, there are endless examples. The hypotenuse of a right triangle depends on its two sides, not necessarily equal. (The function of course involves three variables, but it is called a two-variable function because it has two independent variables.) If z is the hypotenuse, we know from the Pythagorean theorem that $z = \sqrt{x^2 + y^2}$. Note that this is not a one-to-one function. Knowing x and y gives z a unique value, but knowing z does not yield unique values for x and y.

Two other familiar examples of a two-variable function, neither one-to-one, are the area of a triangle as a function of its altitude and base, and the area of a right circular cylinder as a function of its radius and height.

Functions of one and two variables are ubiquitous in physics. The period of a pendulum is a function of its length. The distance covered by a dropped stone and its velocity are each functions of the elapsed time since it was dropped. Atmospheric pressure is a function of altitude. A bullet's energy is a two-variable function dependent on its mass and velocity. The electrical resistance of a wire depends on the length of the wire and the diameter of its circular cross section.

Functions can have any number of independent variables. A simple instance of a three-variable function is the volume of a rectangular room. It is dependent on the room's two sides and height. The volume of a four-dimensional hyper-room is a function of four variables.

A beginning student of calculus must be familiar with how equations with two variables can be modeled by curves on the Cartesian plane. (The plane is named after the French mathematician and philosopher René Descartes who invented it.) Values of the independent variable are represented by points along the horizontal x axis. Values of the dependent variable are represented by points along the vertical y axis. Points on the plane signify an ordered pair of x and y numbers. If a function is linear—that is, if it has one form $y = ax + b$—the curve representing the ordered pairs is a straight line. If the function does not have the form $ax + b$ the curve is not a straight line.

Figure 1 is a Cartesian graph of $y = x^2$. The curve is a parabola. Points along each axis represent real numbers (rational and irra-

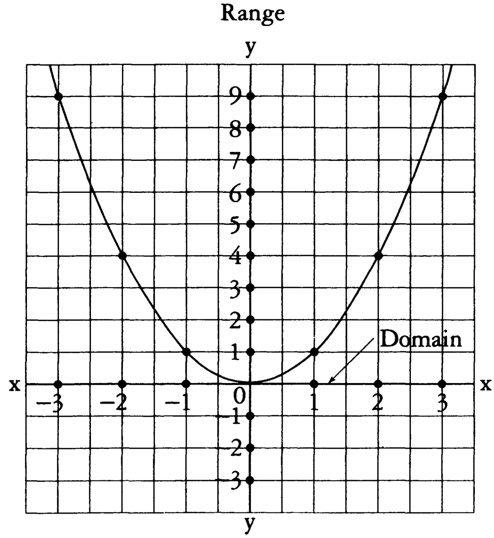

FIG. 1. $y = x^2$ or $f(x) = x^2$
Note that the scales are different on the two axes.

tional), positive on the right side of the x axis, negative on the left; positive at the top of the y axis, negative at the bottom. The graph's origin point, where the axes intersect, represents zero. If x is the side of a square, we assume it is neither zero nor negative, so the relevant curve would be only the right side of the parabola. Assume the square's side is 3. Move vertically up from 3 on the x axis to the curve, then go left to the y axis where you find that the square of 3 is 9. (I apologize to readers for whom all this is old hat.)

If a function involves three independent variables, the Cartesian graph must be extended to a three-dimensional space with axes x, y, and z. I once heard about a professor, whose name I no longer recall, who liked to dramatize this space to his students by running back and forth while he exclaimed "This is the x axis!" He then ran up and down the center aisle shouting "This is the y axis!", and finally hopped up and down while shouting "This is the z axis!" Functions of more than three variables require a Cartesian space with more than three axes. Unfortunately, a professor cannot dramatize axes higher than three by running or jumping.

Note the labels "domain" and "range" in Figure 1. In recent decades it has become fashionable to generalize the definition of function. Values that can be taken by the independent variable are called the variable's *domain*. Values that can be taken by the dependent variable are called the *range*. On the Cartesian plane the domain consists of numbers along the horizontal (x) axis. The range consists of numbers along the vertical (y) axis.

Domains and ranges can be infinite sets, such as the set of real numbers, or the set of integers; or either one can be a finite set such as a portion of real numbers. The numbers on a thermometer, for instance, represent a finite interval of real numbers. If used to measure the temperature of water, the numbers represent an interval between the temperatures at which water freezes and boils. Here the height of the mercury column relative to the water's temperature is a one-to-one function of one variable.

In modern set theory this way of defining a function can be extended to completely arbitrary sets of numbers for a function that is described not by an equation but by a set of rules. The simplest way to specify the rules is by a table. For example, the table in Figure 2 shows a set of arbitrary numbers that constitute the domain on the left. The corresponding set of arbitrary numbers in the range is on the right. The rules that govern this function are indicated by arrows. These arrows show that every number in the domain correlates to a single number on the right. As you can see, more than one number on the left can lead to the same number on the right, but not vice versa. Another example of such a function is shown in Figure 3, along with its graph, consisting of 6 isolated points in the plane.

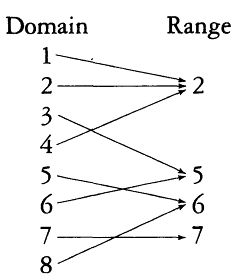

FIG. 2. An arbitrary function.

Because every number on the left leads to exactly one number on the right, we can say that the numbers on the right are a function of those on the left. Some writers call the numbers on the right "images" of those on the left. The arrows are said to furnish a "mapping" of domain to range. Some call the arrows "correspondence rules" that define the function.

For most of the functions encountered in calculus, the domain consists of a single in-

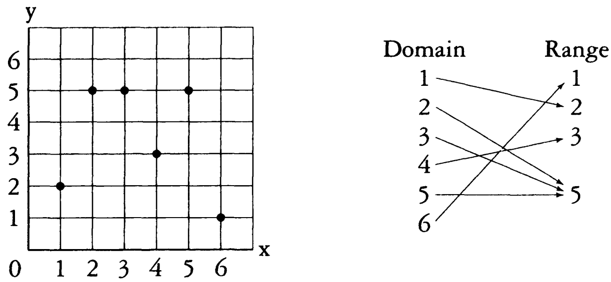

FIG. 3. How another arbitrary discrete function of integers is graphed.

terval of real numbers. The domain might be the entire x axis, as it is for the function $y = x^2$. Or it might be an interval that's bounded; for example, the domain of $y = \arcsin x$ consists of all x such that $-1 \leq x \leq 1$. Or it might be bounded on one side and unbounded on the other; for example, the domain of $y = \sqrt{x}$ consists of all $x \geq 0$. We call such a function "continuous" if its graph can be drawn without lifting the pencil from the paper, and "discontinuous" otherwise. (The complete definition of continuity, which is also applicable to functions with more complicated domains, is beyond the scope of this book.)

For example, the three functions just mentioned are all continuous. Figure 4 shows an example of a discontinuous function. Its domain consists of all real numbers, but its graph has infinitely many pieces that aren't connected to each other. In this book we will be concerned almost entirely with continuous functions.

Note that if a vertical line from the x axis intersects more than one point on a curve, the curve cannot represent a function because it maps an x number to more than one y number. Figure 5 is a graph that clearly is not a function because vertical lines, such as the one shown dotted, intersect the graph at three spots. (It should be noted that Thompson did not use the modern definition of "function." For example the graph shown in Figure 30 of Chapter XI fails this vertical line test, but Thompson considers it a function.)

In this generalized definition of function, a one-variable function is any set of ordered pairs of numbers such that every num-

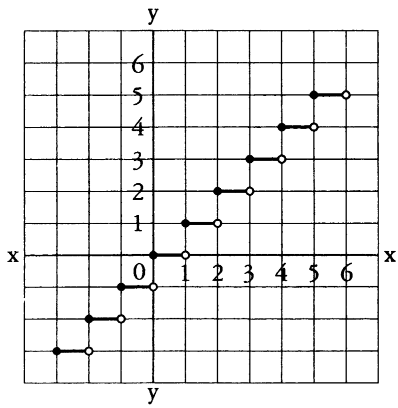

FIG. 4. This function is called the greatest integer function because it maps each real number (on the x axis) to the largest integer on the y axis that is equal to or less than the real number.

ber in one set is paired with exactly one number of the other set. Put differently, in the ordered pairs no x number can be repeated though a y number can be.

In this broad way of viewing functions, the arbitrary combination of a safe or the sequence of buttons to be pushed to open a door, are functions of counting numbers. To open a safe you must turn the knob back and forth to a random set of integers. If the safe's combination is, say, 2-19-3-2-19, then those numbers are a function of 1,2,3,4,5. They represent the order in which numbers must be taken to open the safe, or the order in which buttons must be pushed to open a door. In a similar way the heights of the tiny "peaks" along a cylinder lock's key are an arbitrary function of positions along the key's length.

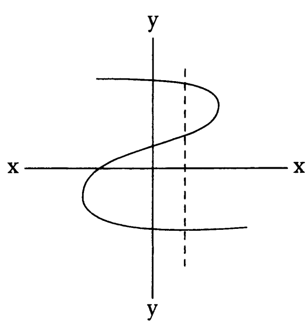

FIG. 5. A graph that does not represent a function.

In recent years mathematicians have widened the notion of function even further to include things that are not numbers. Indeed, they can be anything at all that are elements of a set. A function is simply the correlation of each element in one set to exactly one element of another set. This leads to all sorts of uses of the word function that seem absurd. If Smith has red hair, Jones has black hair, and Robinson's hair is white, the hair color is a function of the three men. Positions of towns on a map are a function of their positions on the earth. The number of toes in a normal family is a function of the number of persons in the family. Different persons can have the same mother, but no person has more than one mother. This allows one to say that mothers are a function of persons. Elephant mothers are a function of elephants, but not grandmothers because an elephant can have two grandmothers. As one mathematician recently put it, functions have been generalized "up to the sky and down into the ground."

A useful way to think of functions in this generalized way is to imagine a black box with input and output openings. Any element in a domain, numbers or otherwise, is put into the box. Out will pop a single element in the range. The machinery inside the box magically provides the correlations by using whatever correspondence rules govern the function. In calculus the inputs and outputs are almost always real numbers, and the machinery in the black box operates on rules provided by equations.

Because the generalized definition of a function leads to bizarre extremes, many educators today, especially those with engineering backgrounds, think it is confusing and unnecessary to introduce such a broad definition of functions to beginning calculus students. Nevertheless, an increasing number of modern calculus textbooks spend many pages on the generalized definition. Their authors believe that defining a function as a mapping of elements from any set to any other set is a strong unifying concept that should be taught to all calculus students.

Opponents of this practice think that calculus should not be concerned with toes, towns, mothers, and elephants. Its domains and ranges should be confined, as they have always been, to real numbers whose functions describe continuous change.

It is a fortunate and astonishing fact that the fundamental laws of our fantastic fidgety universe are based on relatively simple equations. If it were otherwise, we surely would know less than we know now about how our universe behaves, and Newton and Leibniz would probably never have invented (or discovered?) calculus.

Preliminary Chapter 2

WHAT IS A LIMIT?

It is possible, though difficult, to understand calculus without a firm grasp on the meaning of a limit. A derivative, the fundamental concept of differential calculus, is a limit. An integral, the fundamental concept of integral calculus, is a limit.

To explain what is meant by a limit, we will be concerned in this chapter only with limits of discrete functions because limits are easier to understand in discrete terms. When you read *Calculus Made Easy* you will learn how the limit concept applies to what are called functions of a continuous variable because their variables have real number values that vary continuously. Functions of discrete variables have variables whose values jump from one value to another. There are also functions of complex variables in which the values are complex numbers—numbers based on the imaginary square root of minus one. Complex variables are outside the scope of Thompson's book.

A sequence is a set of numbers in some order. The numbers don't have to be different and they need not be integers. Consider the sequence 1,2,3,4,. . . . This is just the positive integers. It is an infinite sequence because it continues without stopping. If it stopped it would be a finite sequence.

If the terms of a finite sequence are added to obtain a finite sum, it is called a series. If a series is infinite, the sum up to any specified term is called a "partial sum." If the partial sums of an infinite series get closer and closer to a number k, so that by continuing the series you can make the sum as close to k as you please, then k is called the limit of the partial sums, or the limit of the infinite series. The terms are said to "converge" on k. If there is no convergence, the series is said to "diverge."

The limit of an infinite series is sometimes called its "sum at infinity," but of course this is not a sum in the usual arithmetical sense when the number of terms is finite. You can't obtain the "sum" of an infinite series by adding because the number of terms to be added is infinite. When we speak of the "sum" of an infinite series, this is just a short way of naming its limit.

An infinite series can converge on its limit in three different ways:

1. The partial sums get ever closer to the limit without actually reaching it, but they never go beyond the limit.
2. The partial sums reach the limit.
3. The partial sums go beyond the limit before they converge.

Let's look at examples of types 1 and 3.

The fifth century B.C. Greek philosopher Zeno of Elia invented several famous paradoxes intended to show that there is something extremely mysterious about motion. One of them imagines a runner going from A to B. He first runs half the distance, then half the remaining distance, then again half the remaining distance, and so on. The distances he runs get smaller and smaller in the halving series $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots\ \frac{1}{2^n}$. Distances from B approach zero as their limit while the distances from A form a series that converges on 1. The runner, of course, models a point moving along a line from A to B. Does the runner ever reach the goal?

It depends.

Assume that after each step in the series the runner pauses to rest for a second. We can model this with a pawn (representing a point) that you push across a table from one edge to the edge opposite. First you push the pawn half the distance, then pause for a second. You push it half the remaining distance and again pause for a second. If this procedure continues, the pawn (point) will get closer and closer to the limit, but will never reach it.

There is an old joke based on this. A mathematics professor places a male student at one side of an empty room and a gorgeous female student at the opposite wall. On command, the boy walks half the distance toward the girl, waits a second, then goes half the rest of the way, and so on, always pausing a second before he cuts the remaining distance in half. The girl says, "Ha ha, you'll never reach me!" The boy replies, "True, but I can get close enough for all practical purposes."

Suppose, now, that instead of waiting a second after each pawn push, the pawn is moved at a steady rate. Assume that the constant speed is such that the pawn goes half the distance in one second, half the remaining distance in half a second, and so on. No pauses. A discrete process has been transformed into a continuous one. In two seconds the pawn has reached the table's far edge. Zeno's runner, if he goes at a steady rate, will reach the goal in a finite period of time. The halving series, modeled in this fashion, converges exactly on the limit.

Zeno's runner leads to a variety of amusing paradoxes involving what are called "infinity machines." A simple example is a lamp that is turned off at the end of one minute, then turned on at the end of half a minute, off after a quarter minute, and so on in an infinite series of ons and offs. The time series converges on two minutes. At the end of two minutes is the lamp on or off? This of course is a thought experiment. It can't be performed with an actual lamp, but can it be answered in the abstract? No, because there *is* no last operation in an infinite series of on and off. It is like asking if the last digit of pi is odd or even.*

An easy way to "see" that the limit of $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots$ is 1 is to mark off the fractional lengths along a number line as Thompson does in his Figure 46. A similar "look-see" proof that the series converges on 1 is shown by the dissected unit square in Figure 6. The

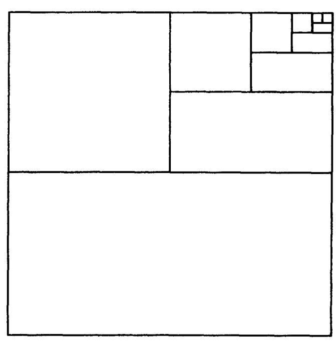

FIG. 6. A two-dimensional "look-see" proof that $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots=1$.

*On infinity machines, see "Alephs and Supertasks," Chapter 4, in my *Wheels, Life, and Other Mathematical Amusements* (W. H. Freeman, 1983), and the references cited in that chapter's bibliography.

partial sums of this series are generated by the discrete function $1 - \frac{1}{2^n}$, where *n* takes the integral values 1,2,3,4,5, ...

We turn now to an infinite series that goes past its limit before finally converging. An example is provided by changing every other sign in the halving series to a minus sign: $\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \ldots$. The partial sums of this "alternating series" are alternately above and below the limit of $\frac{1}{3}$. The difference from $\frac{1}{3}$ can be made as small as you please, but every other partial sum is larger than the limit.

As an infinite series approaches but never reaches its limit, the differences between a partial sum and the limit get closer and closer to zero. Indeed they get so close that you can assume they *are* zero and therefore, as Thompson likes to say, they can be "thrown away." In early books on calculus, *terms* said to become infinitely close to zero were called "infinitesimals." Clearly there is something spooky about numbers living in a neverland that is infinitely close to zero, yet somehow not zero. In the halving series, for example, the fractions approaching zero never become infinitesimals because they always remain a finite portion of 1. Infinitesimals are an infinitely small part of 1. They are smaller than any finite fraction you can name, yet never zero. Are they legitimate mathematical entities, or should they be banished from mathematics?

The most outspoken opponent of infinitesimals was the eighteenth-century British philosopher Bishop George Berkeley who attacked them in a 1734 book titled *The Analyst, Or a Discourse Addressed To an Infidel Mathematician.* The infidel was the astronomer Edmond Halley, for whom Halley's comet is named, and the man who persuaded Newton to publish his famous *Principia.*

Here are some of Bishop Berkeley's complaints about infinitesimals. ("Fluxion" was Newton's term for a derivative.)

> And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?

> And of the aforesaid fluxions there be other fluxions, which fluxions of fluxions are called second fluxions. And the fluxions of these second fluxions are called third fluxions: and so on, fourth, fifth, sixth, etc., *ad infinitum.* Now, as our Sense is strained and puzzled with the perception of objects extremely minute, even so the Imagination, which faculty derives from sense, is very much strained and puzzled to frame clear ideas of the least particle of time, or the least increment generated therein: and much more to comprehend the moments, or those increments of the flowing quantities in *status nascenti,* in their first origin or beginning to exist, before they become finite particles. And it seems still more difficult to conceive the abstracted velocities of such nascent imperfect entities. But the velocities of the velocities, the second, third, fourth, and fifth velocities, etc., exceed, if I mistake not, all human understanding. The further the mind analyseth and pursueth these fugitive ideas the more it is lost and bewildered; the objects, at first fleeting and minute, soon vanishing out of sight. Certainly, in any sense, a second or third fluxion seems an obscure Mystery. The incipient celerity of an incipient celerity, the nascent augment of a nascent augment, i. e. of a thing which hath no magnitude; take it in what light you please, the clear conception of it will, if I mistake not, be found impossible; whether it be so or no I appeal to the trial of every thinking reader. And if a second fluxion be inconceivable, what are we to think of third, fourth, fifth fluxions, and so on without end.
>
> He who can digest a second or third fluxion, a second or third difference, need not, methinks, be squeamish about any point in Divinity.

Johann Bernoulli, a Swiss mathematician who did pioneering work in developing calculus, expressed the paradox of infinitesimals crisply. They are so tiny, he said, that "if a quantity is increased or decreased by an infinitesimal, then that quantity is neither increased nor decreased."

For two centuries most mathematicians agreed with Berkeley and refused to use the term. You won't find it in *Calculus Made*

Easy. Bertrand Russell, in *Principles of Mathematics* (1903, Chapters 39 and 40) has a vigorous attack on infinitesimals. He calls them "mathematically useless," "unnecessary, erroneous, and self-contradictory." As late as 1941 the noted mathematician Richard Courant wrote: "Infinitely small quantities are now definitely and dishonorably discarded." Like Russell and others, he believed that calculus should replace infinitesimals by the concept of limits.

Charles Peirce (1839-1914), America's great mathematician and philosopher, and friend of William James, strongly disagreed. He was almost alone in his day in siding with Leibniz, who believed that infinitesimals were as real and as legitimate as imaginary numbers. Here are some typical remarks by Peirce that I found by checking "infinitesimal" in the indexes of the volumes that make up Peirce's *Collected Papers* and his *New Elements of Mathematics.*

> Infinitesimals may exist and be highly important for philosophy, as I believe they are.
>
> The doctrine of infinitesimals is far simpler than the doctrine of limits.
>
> Is it consistent . . . freely to admit of imaginaries while rejecting infinitesimals as inconceivable?

> Infinitesimals, in the strict and literal sense, are perfectly intelligible, contrary to the teaching of the great body of modern textbooks on the calculus.
>
> There is nothing contradictory about the idea of such quantities. . . . As a mathematician, I prefer the method of infinitesimals to that of limits, as far easier and less infested with snares.

Peirce would have been delighted had he lived to see the work of Abraham Robinson, of Yale University. In 1960, to the vast surprise of mathematicians everywhere, Robinson found a way to reintroduce Leibniz's infinitesimals as legitimate, precisely defined mathematical entities! His way of using them in calculus is known as "nonstandard analysis." (Analysis is a term applied to calculus and all higher mathematics that use calculus.) Nonstandard analysis has produced simpler solutions than standard analysis to many calculus problems, and of course it is closer to an intuitive way of interpreting infinite converging series. Robinson's achievement is too difficult to go into here, but you will find a good introduction to it in "Nonstandard Analysis," by Martin Davis and Reuben Hersh, in *Scientific American,* June 1972.

Mathematician and science fiction writer Rudy Rucker, in his book *Infinity and the Mind* (1982) vigorously defends infinitesimals:

> So great is the average person's fear of infinity that to this day calculus all over the world is being taught as a study of *limit processes* instead of what it really is: *infinitesimal analysis.*
>
> As someone who has spent a good portion of his adult life teaching calculus courses for a living, I can tell you how weary one gets of trying to explain the complex and fiddling theory of limits to wave after wave of uncomprehending freshmen. . . .
>
> But there is hope for a brighter future. Robinson's investigations of the hyperreal numbers have put infinitesimals on a logically unimpeachable basis, and here and there calculus texts based on infinitesimals have appeared.

Which is preferable? To talk about quantities so infinitely small that you can, as Thompson says, "throw them away," or to talk of values approaching a limit? Debate over the infinitesimal versus the limit language goes nowhere because they are two ways of saying the same thing. It's like choosing between calling a triangle a polygon with three sides or a polygon with three angles. Calculations in differentiating or integrating are exactly the same regardless of your preference for how to describe what you are doing. Now that infinitesimals have become respectable again, thanks to nonstandard analysis, you needn't hesitate, if you like, to use the term.

You might suppose that if the terms of an infinite series get smaller and smaller, the series must converge. This is far from true. The most famous example is $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots$. Known as the "harmonic series," it has countless applications in physics as well as in mathematics. Although its fractions get progressively smaller, converging on zero, its partial sums grow without limit! The growth is infuriatingly slow. After a hundred terms the partial sum is only a bit higher than 5. To reach a sum of 100 requires more than 10^{43} terms!

If we eliminate all terms in the harmonic series that have even denominators, will it converge? Amazingly, it will not, though its rate of growth is much slower. If we eliminate from the series all terms whose denominators contain a specific digit one or more times, the series *will* then converge. The following table gives to two decimal places the limit for each omitted digit:

Omitted digit	*Sum*
1	16.18
2	19.26
3	20.57
4	21.33
5	21.83
6	22.21
7	22.49
8	22.73
9	22.92
0	23.10

Limits of infinite series can be expressed by unending decimal fractions. For example, .33333. . . . is the limit of the series $\frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \ldots = \frac{1}{3}$. Incidently, there is a ridiculously easy way to determine the integral limit of any repeating decimal. The trick is to divide the repetend (the repeated sequence of digits) by a number consisting of the same number of nines as there are digits in the repetend. Thus .3333. . . reduces to $\frac{3}{9} = \frac{1}{3}$. If the repeating decimal is, say, .123123123. . . . the limit is $\frac{123}{999}$ which reduces to $\frac{41}{333}$.

Irrational numbers such as irrational roots, and transcendental numbers such as pi and *e*, are limits of many infinite series. Pi, for example, is the limit of such highly patterned series as: $\frac{4}{1} - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \ldots$. The number *e* (you will encounter it in Thompson's Chapter 14) is the limit of $1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots$.

Although Archimedes did not know calculus, he anticipated integration by calculating pi as the limit of the perimeters of regular polygons as their number of sides increases. In the language of infinitesimals, a circle can be viewed as the perimeter of a regular polygon with an infinity of sides, its perimeter consisting of an infinity of straight line segments each of infinitesimal length.

Many ingenious techniques have been found for determining if an infinite series converges or diverges, as well as ways, sometimes not easy, of finding the limit. If the terms of a series decrease in a geometric progression (each term is the same fraction of the preceding one) finding the limit is easy. Here is how it works on the halving series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\ldots$ Let x equal the entire series. Multiply each side of the equation by 2:

$$2x = 2 + \tfrac{2}{2} + \tfrac{2}{4} + \tfrac{2}{8} + \tfrac{2}{16} + \ldots$$

Reduce the terms:

$$2x = 2 + 1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \ldots.$$

Note that the series beyond 2 is the same as the original halving series which we took as x. This enables us to substitute x for the sequence and write $2x = 2 + x$. Rearranging terms to $2x - x = 2$ gives x, the limit of the series, a value of 2.

The same trick will show that $\frac{1}{2}$ is the limit of $\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \frac{1}{81} + \ldots$; it works on any series in which terms decrease in geometric progression.

Bouncing ball problems are common in the literature on limits. They assume that an ideally elastic ball is dropped a specified distance to a hard floor. After each bounce it rises a constant fraction of the previous height. Here is a typical example.

The ball is dropped from a height of four feet. Each bounce takes it to $\frac{3}{4}$ the previous height. In practice, of course, a rubber ball bounces only a finite number of times, but the idealized ball bounces an infinite number of times. The rises approach zero as a limit, but because the times of each bounce also approach a limit of zero, the ball (like Zeno's runner) finally reaches the limit. After an infinity of bounces, it comes to rest after a finite period of time. When the ball ceases to bounce, how far has it traveled?

We can solve this problem by using the same trick used on the halving series. Ignoring for a moment the initial drop of four feet, the ball will rise three feet then fall three feet for a total of six feet. After that, each bounce (rise plus fall) is three-fourths the previous bounce. Letting x be the total distance the ball travels after the first drop of four feet, we write the equation:

$$x = 6 + \tfrac{18}{4} + \tfrac{54}{16} + \tfrac{162}{64} + \tfrac{486}{256} + \ldots$$

Reducing the fractions:

$$x = 6 + \tfrac{9}{2} + \tfrac{27}{8} + \tfrac{81}{32} + \tfrac{243}{128} + \ldots$$

Because each term is $\frac{4}{3}$ its following term, we multiply each side by $\frac{4}{3}$ to get:

$$\tfrac{4x}{3} = 8 + 6 + \tfrac{9}{2} + \tfrac{27}{8} + \tfrac{81}{32} + \ldots.$$

Observe that after 8 the sequence is the same as x, so we can substitute x for it:

$$\tfrac{4x}{3} = 8 + x$$

$$4x = 24 + 3x$$

$$x = 24$$

This is the distance the ball bounces after the initial drop of 4 feet. The total distance traveled by the ball is $24 + 4 = 28$ feet.

Sam Loyd, America's great puzzlemaker, in his *Cyclopedia of Puzzles* (page 23), and his British counterpart Henry Ernest Dudeney, in *Puzzles and Curious Problems* (Problem 223), each give the following ball bounce problem. A ball is dropped 179 feet from the Tower of Pisa. Each bounce is one-tenth the height of the previous bounce. How far does the ball travel after an infinity of bounces before it finally comes to rest? (See Figure 7.)

We can solve this by the trick used before, but because each fraction is one-tenth the previous one, we can find the answer by an even faster method.

After the initial drop of 179 feet, the height of the first bounce is 17.9. Succeeding bounces have heights of 1.79, .179, .0179, and so on, with the decimal point moving one position left after each bounce. Adding these heights gives a total of 19.8888. . . . We now double this distance to obtain the up plus down distances for each bounce to get 39.7777. . . . Finally, we add the

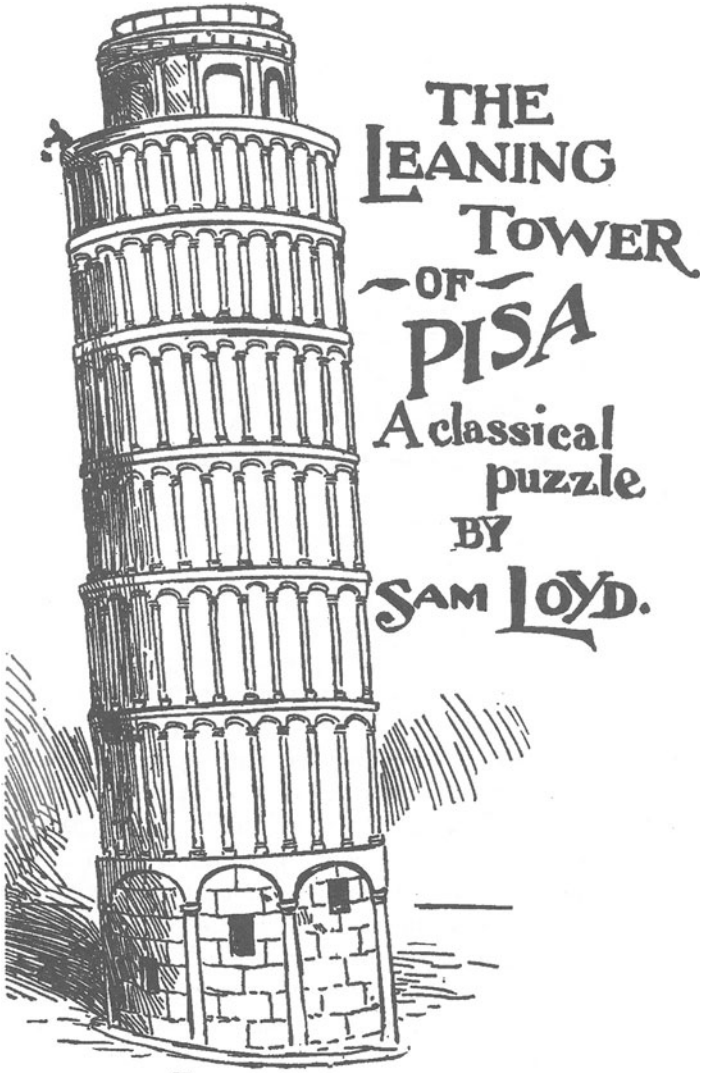

FIG. 7. Sam Loyd's bouncing ball puzzle.

initial drop of 179 feet to obtain the total distance the ball travels: 218.7777. . . , or exactly 218 and $\frac{7}{9}$ feet.

Converging series that do not decrease by a geometric progression often can be solved by other clever methods. Here is an interesting example.

$$x = 1 + \tfrac{3}{2} + \tfrac{5}{4} + \tfrac{7}{8} + \tfrac{9}{16} + \tfrac{11}{32} + \ldots.$$

Note that the numerators are odd numbers in sequence, and the denominators are a doubling series. Here is a simple way to find the limit.

First divide each term by 2:

$$\tfrac{x}{2} = \tfrac{1}{2} + \tfrac{3}{4} + \tfrac{5}{8} + \tfrac{7}{16} + \ldots.$$

Subtract this sequence from the original sequence:

$$\begin{array}{l} x = 1 + \tfrac{3}{2} + \tfrac{5}{4} + \tfrac{7}{8} + \tfrac{9}{16} + \ldots \\ \underline{\tfrac{x}{2} = \phantom{1 +{}} \tfrac{1}{2} + \tfrac{3}{4} + \tfrac{5}{8} + \tfrac{7}{16} + \ldots} \\ \tfrac{x}{2} = 1 + \left[1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \ldots.\right] \end{array}$$

Observe that after the 1 inside the brackets, the sequence that follows is our old friend the halving series which we know converges on 1. Adding 2 to the initial 1 gives the series a limit of 3. Since 3 is half of x, x must be 6, the limit of the original series.

Thompson does not spend much time on series and their limits. I have done so in this chapter for two reasons: they are the best way to become comfortable with the limit concept, and modern calculus textbooks now usually include chapters on infinite series and their usefulness in many aspects of calculus.

Preliminary Chapter 3

WHAT IS A DERIVATIVE?

In Chapter 3 Thompson makes crystal clear what a derivative is, and how to calculate it. However, it seemed to me useful to make a few introductory remarks about derivatives that may make Thompson's chapter even easier to understand.

Let's start with Zeno's runner. Assume that he runs ten meters per second on a path from zero to 100 meters. The independent variable is time, represented by the x axis of a Cartesian graph. The dependent variable y is the runner's distance from his starting spot. It is represented on the y axis. Because the function is linear, the runner's motion graphs as an upward tilted straight line from zero, the graph's origin, to the point that is ten seconds on the time axis and 100 meters on the distance axis. (Figure 8) If by distance we mean distance *from* the goal, the line on the graph tilts the other way (Figure 9).

Given any point in time, how fast is the runner moving? Because we are dealing with a simple linear function we don't need calculus to tell us that at every instant he is going ten meters per second. The function's equation is $y = 10x$. Note that the slope of the line on the graph, as measured by the height in meters at any point divided by the elapsed time in seconds at any point, is 10. At each instant the runner has gone in meters ten times the elapsed number of seconds. His instantaneous speed throughout the run clearly is ten meters per second.

Consider any moment of time along the x axis, then go vertically up the graph to the distance traveled in meters. You will find that the distance is always ten times the elapsed time. As you will learn when you read *Calculus Made Easy,* the derivative of a

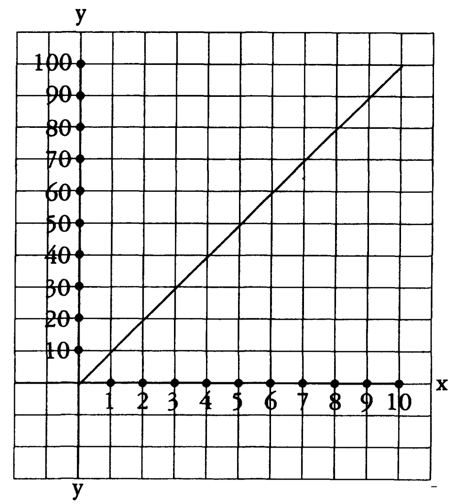

FIG. 8. Graph of Zeno's runner.
The x axis is time, the y axis is the distance from the start of the run.

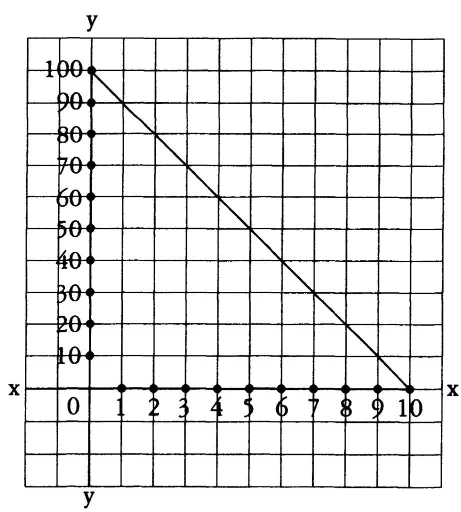

FIG. 9. Graph of Zeno's runner showing distance from goal. The equation is $y = 10(10 - x)$.

function is simply another function that describes the rate at which a dependent variable changes with respect to the rate at which the independent variable changes. In this case the runner's speed never changes, so the derivative of $y = 10x$ is simply the number 10. It tells you two things: (1) that at any time the runner's speed is ten meters per second, and (2) that at any point on the line that graphs this function, the slope of the line is 10. This generalizes to all linear functions in which the variable y changes with respect to variable x at a constant rate. If a function is $y = ax$, its derivative is simply the constant a.

As I said, you don't need calculus to tell you all this, but it is good to know that calculating derivatives gives the correct result even when functions are linear.

An even simpler case of a derivative, too obvious to require any thought, let alone demanding calculus, is the case of a runner who stands perfectly still. Let's say he stops running after going ten meters. The function is $y = 10$. The graph becomes a horizontal straight line as shown in Figure 10. Its slope is zero which is the same as saying that the rate at which the runner's distance from the start changes, relative to changes of time, is zero. The function's derivative is zero. Even in this extreme case it is comforting to know that calculus still applies. In general, the derivative of any constant is zero.

Calculus ceases to be trivial when functions are nonlinear. Consider the simple nonlinear function $y = x^2$, which Thompson uses to open his chapter on derivatives. Let's see how it applies to the growth of a square, the simplest geometrical interpretation of this function.

Imagine a monster living on Flatland, a plane of two dimensions. It is born a square of side 1 and area 1, then grows at a steady rate. We wish to know, at any instant of time, how fast its area grows with respect to the growth of its side.

The monster's area, of course, is the square of its side, so the function we have to consider is

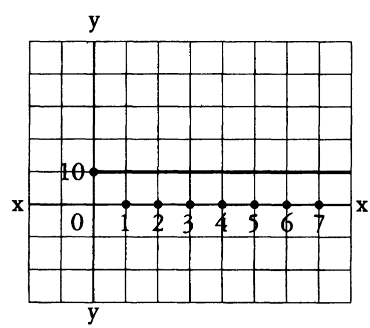

FIG. 10. Graph of a runner who stands still at a distance of ten units from the start.

$y = x^2$, where y is the area and x the side. (It graphs as the parabola shown in Figure 1 of the first preliminary chapter.) As you will learn from Thompson, the function's derivative is $2x$. What does this tell us? It tells us that at any given moment the monster's area is growing at a rate that is $2x$ times as fast as its side is lengthening.

For example, let's say the monster's side is growing at a rate of 3 units per second. Starting with a side of one unit, at the end of ten seconds its side will have reached 31 units. The value of x at this point is 31. The derivative says that when the monster's side is 31, its area is increasing with respect to its side at a rate of $2x$, or $2 \times 31 = 62$ units. When the square reaches 100 on its side, its area will be increasing with respect to its side by $2 \times 100 = 200$ units.

These figures express the rate of the square's growth with respect to its side. For the square's growth rate with respect to *time* we have to multiply these values by 3. Thus when the square has a side of 31 (after ten seconds), it is growing at a rate of $3 \times 2 \times 31 = 186$ square units per second. When the side is 100, its growth rate per second is $3 \times 2 \times 100 = 600$.

Suppose the monster is a cube of edge x which increases at a steady rate of 2 units per second. The cube's volume, y, is x^3. The derivative of the function $y = x^3$ is $3x^2$. This tells you that the cube's volume in cubical units grows $3x^2$ times as fast as its edge grows. Thus when the cube's side reaches, say, 10, the value of x, its volume, is growing $3 \times 10^2 = 300$ square inches as fast as its side. Its growth rate per second is $2 \times 3 \times 10^2 = 600$.

Although Thompson avoids defining a derivative as the limit of a sequence of ratios, this clearly is the case. Suppose, for instance, that our growing square has sides that increase at one unit per second. We can tabulate the area's growth at times slightly larger than 2 seconds as follows:

Time	*Side*	*Area*
2	3	9
2.1	3.1	9.61
2.01	3.01	9.0601
2.001	3.001	9.006001

The average rate of growth from time 2 to time 2.1 is:

$$\frac{9.61 - 9}{2.1 - 2} = 6.1$$

And from time 2 to time 2.01:

$$\frac{9.0601 - 9}{2.01 - 2} = 6.01$$

And from time 2 to time 2.001:

$$\frac{9.006001 - 9}{2.001 - 2} = 6.001$$

The averages obviously approach a limit of 6. Thus the derivative of the area with respect to time is the limit of an infinite sequence of ratios that converge on 6. Put simply, a derivative is the rate at which a function's dependent variable grows with respect to the growth rate of the independent variable. In geometrical terms, it determines the exact slope of the tangent to a function's curve at any specified point along the curve. This equivalence of the algebraic and geometrical definitions of a derivative is one of the most beautiful aspects of calculus.

I hope this and the previous two preliminary chapters will help prepare you for understanding *Calculus Made Easy.*

CALCULUS MADE EASY

What one fool can do, another can.

—*Ancient Simian proverb*

PUBLISHER'S NOTE ON THE THIRD EDITION

Only once in its long and useful life in 1919, has this book been enlarged and revised. But in twenty-six years much progress can be made, and the methods of 1919 are not likely to be the same as those of 1945. If, therefore, any book is to maintain its usefulness, it is essential that it should be overhauled occasionally so that it may be brought up-to-date where possible, to keep pace with the forward march of scientific development.

For the new edition the book has been reset, and the diagrams modernised. Mr. F. G. W. Brown has been good enough to revise the whole of the book, but he has taken great care not to interfere with the original plan. Thus teachers and students will still recognise their well-known guide to the intricacies of the calculus. While the changes made are not of a major kind, yet their significance may not be inconsiderable. There seems no reason now, even if one ever existed, for excluding from the scope of the text those intensely practical functions, known as the hyperbolic sine, cosine and tangent, whose applications to the methods of integration are so potent and manifold. These have, accordingly, been introduced and applied, with the result that some of the long cumbersome methods of integrating have been displaced, just as a ray of sunshine dispels an obstructing cloud.

The introduction, too, of the very practical integrals:

$$\int e^{pt} \sin kt \cdot dt \quad \text{and} \quad \int e^{pt} \cos kt \cdot dt$$

has eliminated some of the more ancient methods of "Finding Solutions" (Chapter XXI). By their application, shorter and more intelligible ones have grown up naturally instead.

In the treatment of substitutions, the whole text has been tidied up in order to render it methodically consistent. A few examples have also been added where space permitted, while the whole of the exercises and their answers have been carefully revised, checked and corrected. Duplicated problems have thus been removed and many hints provided in the answers adapted to the newer and more modern methods introduced.

It must, however, be emphatically stated that the plan of the original author remains unchanged; even in its more modern form, the book still remains a monument to the skill and the courage of the late Professor Silvanus P. Thompson. All that the present reviser has attempted is to revitalize the usefulness of the work by adapting its distinctive utilitarian bias more closely in relation to present-day requirements.

PROLOGUE

Considering how many fools can calculate, it is surprising that it should be thought either a difficult or a tedious task for any other fool to learn how to master the same tricks.

Some calculus-tricks are quite easy. Some are enormously difficult. The fools who write the text-books of advanced mathematics—and they are mostly clever fools—seldom take the trouble to show you how easy the easy calculations are. On the contrary, they seem to desire to impress you with their tremendous cleverness by going about it in the most difficult way.

Being myself a remarkably stupid fellow, I have had to unteach myself the difficulties, and now beg to present to my fellow fools the parts that are not hard. Master these thoroughly, and the rest will follow. What one fool can do, another can.

COMMON GREEK LETTERS USED AS SYMBOLS

Capital	*Small*	*English Name*	*Capital*	*Small*	*English Name*
A	α	Alpha	Λ	λ	Lambda
B	β	Beta	M	μ	Mu
Γ	γ	Gamma	Ξ	ξ	Xi
Δ	δ	Delta	Π	π	Pi
E	ϵ	Epsilon	P	ρ	Rho
H	η	Eta	Σ	σ	Sigma
Θ	θ	Thēta	Φ	ϕ	Phi
K	κ	Kappa	Ω	ω	Omega

Chapter I

TO DELIVER YOU FROM THE PRELIMINARY TERRORS

The preliminary terror, which chokes off most high school students from even attempting to learn how to calculate, can be abolished once for all by simply stating what is the meaning—in common-sense terms—of the two principal symbols that are used in calculating.

These dreadful symbols are:

(1) d which merely means "a little bit of".

Thus dx means a little bit of x; or du means a little bit of u. Ordinary mathematicians think it more polite to say "an element of", instead of "a little bit of". Just as you please. But you will find that these little bits (or elements) may be considered to be infinitely small.

(2) $\int$ which is merely a long S, and may be called (if you like) "the sum of".

Thus $\int dx$ means the sum of all the little bits of x; or $\int dt$ means the sum of all the little bits of t. Ordinary mathematicians call this symbol "the integral of". Now any fool can see that if x is considered as made up of a lot of little bits, each of which is called dx, if you add them all up together you get the sum of all the dx's (which is the same thing as the whole of x). The word "integral" simply means "the whole". If you think of the duration of time for one hour, you may (if you like) think of it as cut up into 3600 little bits called seconds.

The whole of the 3600 little bits added up together make one hour.

When you see an expression that begins with this terrifying symbol, you will henceforth know that it is put there merely to give you instructions that you are now to perform the operation (if you can) of totalling up all the little bits that are indicated by the symbols that follow.

That's all.

Chapter II

ON DIFFERENT DEGREES OF SMALLNESS

We shall find that in our processes of calculation we have to deal with small quantities of various degrees of smallness.

We shall have also to learn under what circumstances we may consider small quantities to be so minute that we may omit them from consideration. Everything depends upon relative minuteness.

Before we fix any rules let us think of some familiar cases. There are 60 minutes in the hour, 24 hours in the day, 7 days in the week. There are therefore 1440 minutes in the day and 10,080 minutes in the week.

Obviously 1 minute is a very small quantity of time compared with a whole week. Indeed, our forefathers considered it small as compared with an hour, and called it "one minute", meaning a minute fraction—namely one sixtieth—of an hour. When they came to require still smaller subdivisions of time, they divided each minute into 60 still smaller parts, which, in Queen Elizabeth's days, they called "second minutes" (*i.e.* small quantities of the second order of minuteness). Nowadays we call these small quantities of the second order of smallness "seconds". But few people know *why* they are so called.

Now if one minute is so small as compared with a whole day, how much smaller by comparison is one second!

Again, think of a hundred dollars compared with a penny: it is worth only a $\frac{1}{100}$ part. A penny is of precious little importance compared with a hundred dollars: it may certainly be regarded as a *small* quantity. But compare a penny with ten thousand dollars: relative to this greater sum, a penny is of no more importance than a hundredth of a penny would be to a hundred dollars. Even a hundred dollars is relatively a negligible quantity in the wealth of a millionaire.

Now if we fix upon any numerical fraction as constituting the proportion which for any purpose we call relatively small, we can easily state other fractions of a higher degree of smallness. Thus if, for the purpose of time, $\frac{1}{60}$ be called a *small* fraction, then $\frac{1}{60}$ of $\frac{1}{60}$ (being a *small* fraction of a *small* fraction) may be regarded as a *small quantity of the second order* of smallness.*

Or, if for any purpose we were to take 1 percent $\left(i.e.\ \frac{1}{100}\right)$ as a *small* fraction then 1 percent of 1 percent $\left(i.e.\ \frac{1}{10{,}000}\right)$ would be a small fraction of the second order of smallness; and $\frac{1}{1{,}000{,}000}$ would be a small fraction of the third order of smallness, being 1 percent of 1 percent of 1 percent.

Lastly, suppose that for some very precise purpose we should regard $\frac{1}{1{,}000{,}000}$ as "small". Thus, if a first-rate chronometer is not to lose or gain more than half a minute in a year, it must keep time with an accuracy of 1 part in 1,051,200. Now if, for such a purpose, we regard $\frac{1}{1{,}000{,}000}$ (or one millionth) as a small quantity, then $\frac{1}{1{,}000{,}000}$ of $\frac{1}{1{,}000{,}000}$, that is, $\frac{1}{1{,}000{,}000{,}000{,}000}$ will be a small quantity of the second order of smallness, and may be utterly disregarded, by comparison.

Then we see that the smaller a small quantity itself is, the more negligible does the corresponding small quantity of the second order become. Hence we know that *in all cases we are justified in neglecting the small quantities of the second—or third* (or higher)—*orders,* if only we take the small quantity of the first order small enough in itself.

But it must be remembered that small quantities, if they occur in our expressions as factors multiplied by some other factor, may become important if the other factor is itself large. Even a penny becomes important if only it is multiplied by a few hundred.

Now in the calculus we write dx for a little bit of x. These things such as dx, and du, and dy, are called "differentials", the differential of x, or of u, or of y, as the case may be. [You *read*

*The mathematicians may talk about the second order of "magnitude" (*i.e.* greatness) when they really mean second order of *smallness.* This is very confusing to beginners.

them as *dee-eks*, or *dee-you*, or *dee-wy*.] If dx be a small bit of x, and relatively small of itself, it does not follow that such quantities as $x \cdot dx$, or $x^2 dx$, or $a^x dx$ are negligible. But $dx \times dx$ would be negligible, being a small quantity of the second order.

A very simple example will serve as illustration. Consider the function $f(x) = x^2$

Let us think of x as a quantity that can grow by a small amount so as to become $x + dx$, where dx is the small increment added by growth. The square of this is $x^2 + 2x \cdot dx + (dx)^2$. The second term is not negligible because it is a first-order quantity; while the third term is of the second order of smallness, being a bit of a bit of x^2. Thus if we took dx to mean numerically, say, $\frac{1}{60}$ of x, then the second term would be $\frac{2}{60}$ of x^2, whereas the third term would be $\frac{1}{3,600}$ of x^2. This last term is clearly less important than the second. But if we go further and take dx to mean only $\frac{1}{1000}$ of x, then the second term will be $\frac{2}{1000}$ of x^2, while the third term will be only $\frac{1}{1,000,000}$ of x^2.

Geometrically this may be depicted as follows: Draw a square (Fig. 1) the side of which we will take to represent x. Now suppose the square to grow by having a bit dx added to its size each way. The enlarged square is made up of the original square x^2, the two rectangles at the top and on the right, each of which is of area $x \cdot dx$ (or together $2x \cdot dx$), and a little square at the top right-hand corner which is $(dx)^2$. In Fig. 2 we have taken dx as quite a big fraction of x—about $\frac{1}{5}$. But suppose we had taken it only $\frac{1}{100}$—about the thickness of an inked line drawn with a fine pen (See Figure 3). Then the little corner square will have an area of only $\frac{1}{10,000}$ of x^2, and be practically invisible. Clearly $(dx)^2$ is negligible if only we consider the increment dx to be itself small enough.

Let us consider a simile.

Suppose a millionaire were to say to his secretary: next week I will give you a small fraction of any money that comes in to me. Suppose that the secretary were to say to his boy: I will give you a small fraction of what I get. Suppose the fraction in each case to be $\frac{1}{100}$ part. Now if Mr. Millionaire received during the next week \$1,000, the secretary would receive \$10 and the boy 1 dime. Ten dollars would be a small quantity com-

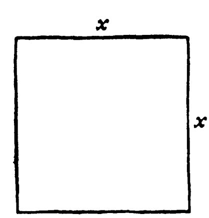

FIG. 1.

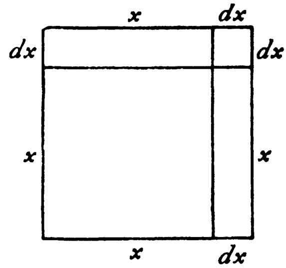

FIG. 2.

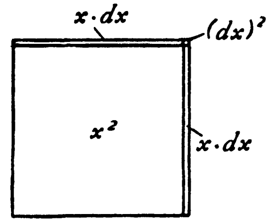

FIG. 3.

pared with \$1,000; but a dime is a small small quantity indeed, of a very secondary order. But what would be the disproportion if the fraction, instead of being $\frac{1}{100}$, had been settled at $\frac{1}{1000}$ part? Then, while Mr. Millionaire got his \$1,000, Mr. Secretary would get only \$1.00, and the boy only a tenth of a penny!

The witty Dean Swift once wrote:

> So, Nat'ralists observe, a Flea
> Hath smaller Fleas that on him prey.
> And these have smaller Fleas to bite 'em.
> And so proceed *ad infinitum.*

An ox might worry about a flea of ordinary size—a small creature of the first order of smallness. But he would probably not trouble himself about a flea's flea; being of the second order of smallness, it would be negligible. Even a gross of fleas' fleas would not be of much account to the ox.

Chapter III

ON RELATIVE GROWINGS

All through the calculus we are dealing with quantities that are growing, and with rates of growth. We classify all quantities into two classes: *constants* and *variables*. Those which we regard as of fixed value, and call *constants*, we generally denote algebraically by letters from the beginning of the alphabet, such as a, b, or c; while those which we consider as capable of growing, or (as mathematicians say) of "varying", we denote by letters from the end of the alphabet, such as x, y, z, u, v, w, or sometimes t.

Moreover, we are usually dealing with more than one variable at once, and thinking of the way in which one variable depends on the other: for instance, we think of the way in which the height reached by a projectile depends on the time of attaining that height. Or, we are asked to consider a rectangle of given area, and to enquire how any increase in the length of it will compel a corresponding decrease in the breadth of it. Or, we think of the way in which any variation in the slope of a ladder will cause the height that it reaches, to vary.

Suppose we have got two such variables that depend on each other. An alteration in one will bring about an alteration in the other, *because* of this dependence. Let us call one of the variables x, and the other that depends on it y.

Suppose we make x to vary, that is to say, we either alter it or imagine it to be altered, by adding to it a bit which we call dx. We are thus causing x to become $x + dx$. Then, because x has been altered, y will have altered also, and will have become $y + dy$. Here the bit dy may be in some cases positive, in others negative; and it won't (except very rarely) be the same size as dx.

Take Two Examples.

(1) Let x and y be respectively the base and the height of a right-angled triangle (Fig. 4), of which the slope of the other side is fixed at 30°. If we suppose this triangle to expand and yet keep its angles the same as at first, then, when the base grows so as to become $x + dx$, the height becomes $y + dy$. Here, increasing x results in an increase of y. The little triangle, the height of which is dy, and the base of which is dx, is similar to the original triangle; and it is obvious that the value of the ratio $\frac{dy}{dx}$ is the same as that of the ratio $\frac{y}{x}$. As the angle is 30° it will be seen that here[1]

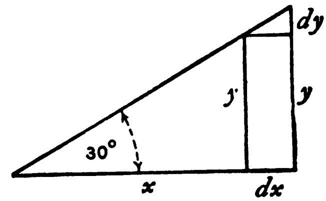

FIG. 4.

$$\frac{dy}{dx} = \frac{1}{1.73\ldots}.$$

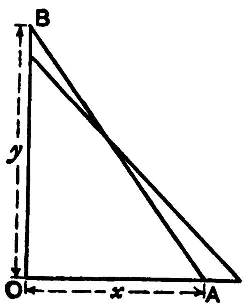

FIG. 5.

(2) Let x represent, in Fig. 5, the horizontal distance, from a wall, of the bottom end of a ladder, AB, of fixed length; and let y be the height it reaches up the wall. Now y clearly depends on x. It is easy to see that, if we pull the bottom end A a bit farther from the wall, the top end B will come down a little lower. Let us state this in scientific language. If we increase x to $x + dx$, then y will become $y - dy$; that is, when x receives a positive increment, the increment which results to y is negative.

Yes, but how much? Suppose the ladder was so long that when the bottom end A was 19 inches from the wall the top end B reached just 15 feet from the ground. Now, if you were to pull the bottom end out 1 inch more, how much would the top end come down? Put it all into inches: $x = 19$ inches, $y = 180$ inches. Now the increment of x which we call dx, is 1 inch: or $x + dx = 20$ inches.

1. The cotangent of 30° is $\sqrt{3} = 1.7320\ldots$. Its reciprocal, $\frac{1}{1.7320}\ldots$, is $.5773\ldots$, the tangent of 30°.—M.G.

How much will y be diminished? The new height will be $y - dy$. If we work out the height by the Pythagorean Theorem, then we shall be able to find how much dy will be. The length of the ladder is

$$\sqrt{(180)^2 + (19)^2} = 181 \text{ inches.}$$

Clearly, then, the new height, which is $y - dy$, will be such that

$$(y - dy)^2 = (181)^2 - (20)^2 = 32761 - 400 = 32361,$$
$$y - dy = \sqrt{32361} = 179.89 \text{ inches.}$$

Now y is 180, so that dy is $180 - 179.89 = 0.11$ inch.

So we see that making dx an increase of 1 inch has resulted in making dy a decrease of 0.11 inch.

And the ratio of dy to dx may be stated thus:

$$\frac{dy}{dx} = \frac{0.11}{1}$$

It is also easy to see that (except in one particular position) dy will be of a different size from dx.

Now right through the differential calculus we are hunting, hunting, hunting for a curious thing, a mere ratio, namely, the proportion which dy bears to dx when both of them are infinitely small.

It should be noted here that we can find this ratio $\frac{dy}{dx}$ only when y and x are related to each other in some way, so that whenever x varies y does vary also. For instance, in the first example just taken, if the base x of the triangle be made longer, the height y of the triangle becomes greater also, and in the second example, if the distance x of the foot of the ladder from the wall be made to increase, the height y reached by the ladder decreases in a corresponding manner, slowly at first, but more and more rapidly as x becomes greater. In these cases the relation between x and y is perfectly definite, it can be expressed mathematically, being $\frac{y}{x} = \tan 30°$ and $x^2 + y^2 = l^2$ (where l is the length of the ladder) respectively, and $\frac{dy}{dx}$ has the meaning we found in each case.

If, while x is, as before, the distance of the foot of the ladder from the wall, y is, instead of the height reached, the horizontal length of the wall, or the number of bricks in it, or the number of years since it was built, any change in x would naturally cause no change whatever in y; in this case $\dfrac{dy}{dx}$ has no meaning whatever, and it is not possible to find an expression for it. Whenever we use differentials dx, dy, dz, etc., the existence of some kind of relation between x, y, z, etc., is implied, and this relation is called a "function" in x, y, z, etc.; the two expressions given above, for instance, namely, $\dfrac{y}{x} = \tan 30°$ and $x^2 + y^2 = l^2$, are functions of x and y. Such expressions contain implicitly (that is, contain without distinctly showing it) the means of expressing either x in terms of y or y in terms of x, and for this reason they are called *implicit functions in* x and y; they can be respectively put into the forms

$$y = x \tan 30° \quad \text{or} \quad x = \frac{y}{\tan 30°}$$

and

$$y = \sqrt{l^2 - x^2} \quad \text{or} \quad x = \sqrt{l^2 - y^2}.$$

These last expressions state explicitly the value of x in terms of y, or of y in terms of x, and they are for this reason called *explicit functions* of x or y. For example $x^2 + 3 = 2y - 7$ is an implicit function of x and y; it may be written $y = \dfrac{x^2 + 10}{2}$ (explicit function of x) or $x = \sqrt{2y - 10}$ (explicit function of y). We see that an explicit function in x, y, z, etc., is simply something the value of which changes when x, y, z, etc., are changing, either one at the time or several together. Because of this, the value of the explicit function is called the *dependent variable,* as it depends on the value of the other variable quantities in the function; these other variables are called the *independent variables* because their value is not determined from the value assumed by the function. For example, if $u = x^2 \sin \theta$, x and θ are the independent variables, and u is the dependent variable.

Sometimes the exact relation between several quantities x, y, z

either is not known or it is not convenient to state it; it is only known, or convenient to state, that there is some sort of relation between these variables, so that one cannot alter either x or y or z singly without affecting the other quantities; the existence of a function in x, y, z is then indicated by the notation $F(x, y, z)$ (implicit function) or by $x = F(y, z)$, $y = F(x, z)$ or $z = F(x, y)$ (explicit function). Sometimes the letter f or ϕ is used instead of F, so that $y = F(x)$, $y = f(x)$ and $y = \phi(x)$ all mean the same thing, namely, that the value of y depends on the value of x in some way which is not stated.

We call the ratio $\frac{dy}{dx}$, "the *differential coefficient* of y with respect to x". It is a solemn scientific name for this very simple thing. But we are not going to be frightened by solemn names, when the things themselves are so easy. Instead of being frightened we will simply pronounce a brief curse on the stupidity of giving long crack-jaw names; and, having relieved out minds, will go on to the simple thing itself, namely the ratio $\frac{dy}{dx}$.[2]

In ordinary algebra which you learned at school, you were always hunting some unknown quantity which you called x or y; or sometimes there were two unknown quantities to be hunted for simultaneously. You have now to learn to go hunting in a new way; the fox being now neither x nor y. Instead of this you have to hunt for this curious cub called $\frac{dy}{dx}$. The process of finding the value of $\frac{dy}{dx}$ is called "differentiating". But, remember, what is wanted is the value of this ratio when both dy and dx are themselves infinitely small. The true value of the derivative is that to which it approximates in the limiting case when each of them is considered as infinitesimally minute.

Let us now learn how to go in quest of $\frac{dy}{dx}$.

2. I have let stand here Thompson's justified criticism of the term "differential coefficient," a term in use when he wrote his book. The term was later replaced by the simpler word "derivative." Henceforth in this book it will be called a derivative.—M.G.

NOTE TO CHAPTER III

How to Read Derivatives

It will never do to fall into the schoolboy error of thinking that dx means d times x, for d is not a factor—it means "an element of" or "a bit of" whatever follows. One reads dx thus: "dee-eks".

In case the reader has no one to guide him in such matters it may here be simply said that one reads derivatives in the following way. The derivative

$\frac{dy}{dx}$ is read *"dee-wy dee-eks"*, or *"dee-wy over dee-eks"*.

So also $\frac{du}{dt}$ is read *"dee-you dee-tee"*.

Second derivatives will be met with later on. They are like this: $\frac{d^2y}{dx^2}$; which is read *"dee-two-wy over dee-eks-squared"*, and it means that the operation of differentiating y with respect to x has been (or has to be) performed twice over.

Another way of indicating that a function has been differentiated is by putting an accent sign to the symbol of the function. Thus if $y = F(x)$, which means that y is some unspecified function of x, we may write $F'(x)$ instead of $\frac{d(F(x))}{dx}$. Similarly, $F''(x)$ will mean that the original function $F(x)$ has been differentiated twice over with respect to x.[3]

3. Newton called a variable a "fluent," and a derivative a "fluxion" because its value flowed or fluctuated continuously. In Chapter 10 Thompson tells how Newton indicated a first derivative by putting a dot over a term, a second derivative by putting two dots, a third derivative by three dots, and so on.

Morris Kline, in his two-volume *Calculus* (1967), is the only mathematician I know of in recent times who adopted Newton's dot notation for derivatives. Physicists, however, often use dot notation to denote differentiation with respect to time.—M.G.

Chapter IV

SIMPLEST CASES

Now let use see how, on first principles, we can differentiate some simple algebraical expression.

Case 1.
Let us begin with the simple expression $y = x^2$.[1] Now remember that the fundamental notion about the calculus is the idea of *growing*. Mathematicians call it *varying*. Now as y and x^2 are equal to one another, it is clear that if x grows, x^2 will also grow. And if x^2 grows, then y will also grow. What we have got to find out is the proportion between the growing of y and the growing of x. In other words, our task is to find out the ratio between dy and dx, or, in brief, to find the value of $\dfrac{dy}{dx}$.

Let x, then, grow a little bit bigger and become $x + dx$; similarly, y will grow a bit bigger and will become $y + dy$. Then, clearly, it will still be true that the enlarged y will be equal to the square of the enlarged x. Writing this down, we have:

$$y + dy = (x + dx)^2$$

Doing the squaring we get:

$$y + dy = x^2 + 2x \cdot dx + (dx)^2$$

What does $(dx)^2$ mean? Remember that dx meant a bit—a little bit—of x. Then $(dx)^2$ will mean a little bit of a little bit of

1. The graph of this equation is a parabola.—M.G.

x^2; that is, as explained above, it is a small quantity of the second order of smallness. It may therefore be discarded as quite negligible in comparison with the other terms. Leaving it out, we then have:

$$y + dy = x^2 + 2x \cdot dx$$

Now $y = x^2$; so let us subtract this from the equation and we have left

$$dy = 2x \cdot dx$$

Dividing across by dx, we find

$$\frac{dy}{dx} = 2x$$

Now *this** is what we set out to find. The ratio of the growing of y to the growing of x is, in the case before us, found to be $2x$.

Numerical Example.
Suppose $x = 100$ and therefore $y = 10{,}000$. Then let x grow till it becomes 101 (that is, let $dx = 1$). Then the enlarged y will be $101 \times 101 = 10{,}201$. But if we agree that we may ignore small quantities of the second order, 1 may be rejected as compared with 10,000; so we may round off the enlarged y to 10,200; y has

N.B.—This ratio $\frac{dy}{dx}$ is the result of differentiating y with respect to x. Differentiating means finding the derivative. Suppose we had some other function of x, as, for example, $u = 7x^2 + 3$. Then if we were told to differentiate this with respect to x, we should have to find $\frac{du}{dx}$, or, what is the same thing, $\frac{d(7x^2+3)}{dx}$. On the other hand, we may have a case in which time was the independent variable, such as this: $y = b + \frac{1}{2}at^2$. Then, if we were told to differentiate it, that means we must find its derivative with respect to t. So that then our business would be to try to find $\frac{dy}{dt}$, that is, to find $\frac{d\left(b + \frac{1}{2}at^2\right)}{dt}$.

grown from 10,000 to 10,200; the bit added on is dy, which is therefore 200.

$\frac{dy}{dx} = \frac{200}{1} = 200$. According to the algebra-working of the previous paragraph, we find $\frac{dy}{dx} = 2x$. And so it is; for $x = 100$ and $2x = 200$.

But, you will say, we neglected a whole unit.

Well, try again, making dx a still smaller bit.

Try $dx = \frac{1}{10}$. Then $x + dx = 100.1$, and

$$(x + dx)^2 = 100.1 \times 100.1 = 10{,}020.01$$

Now the last figure 1 is only one-millionth part of the 10,000, and is utterly negligible; so we may take 10,020 without the little decimal at the end.[2] And this makes $dy = 20$; and $\frac{dy}{dx} = \frac{20}{0.1} = 200$, which is still the same as $2x$.

2. Many writers of calculus texts prefer to use the Greek letter delta, Δ, in place of d to stand for an increment that is small enough to be taken as zero. A derivative is defined as:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

This expresses the limit when Δx diminishes to zero. For example, if $f(x) = 2$, the formula becomes:

$$\frac{\Delta y}{\Delta x} = \frac{2 - 2}{\Delta x} = \frac{0}{\Delta x}$$

Since Δx goes to zero, the derivative of 2 is zero, and its graph is a horizontal line.

If $f(x) = 2x$, the formula gives $2\Delta x/\Delta x$. Because Δx goes to zero, the derivative of $2x$ is 2, and the function graphs as a straight line sloping upward.

Thompson does not use the Δ notation. Indeed, he avoids altogether the notion of limits. But no harm is done. It is easy to translate Thompson's technique of "exhausting" ever decreasing increments to the point where they can be "thrown away" into today's way of defining derivatives as limits.—M.G.

Case 2.
Try differentiating $y = x^3$ in the same way.

We let y grow to $y + dy$, while x grows to $x + dx$.

Then we have

$$y + dy = (x + dx)^3$$

Doing the cubing we obtain

$$y + dy = x^3 + 3x^2 \cdot dx + 3x(dx)^2 + (dx)^3$$

Now we know that we may neglect small quantities of the second and third orders; since, when dy and dx are both made infinitely small, $(dx)^2$ and $(dx)^3$ will become infinitely smaller by comparison. So, regarding them as negligible, we have left:

$$y + dy = x^3 + 3x^2 \cdot dx$$

But $y = x^3$; and, subtracting this, we have:

$$dy = 3x^2 \cdot dx$$

and $$\frac{dy}{dx} = 3x^2$$

Case 3.
Try differentiating $y = x^4$. Starting as before by letting both y and x grow a bit, we have:

$$y + dy = (x + dx)^4$$

Working out the raising to the fourth power, we get

$$y + dy = x^4 + 4x^3dx + 6x^2(dx)^2 + 4x(dx)^3 + (dx)^4$$

Then, striking out the terms containing all the higher powers of dx, as being negligible by comparison, we have

$$y + dy = x^4 + 4x^3dx$$

Subtracting the original $y = x^4$, we have left

$$dy = 4x^3dx, \quad \text{and} \quad \frac{dy}{dx} = 4x^3$$

Now all these cases are quite easy. Let us collect the results to see if we can infer any general rule. Put them in two columns, the values of y in one and the corresponding values found for $\frac{dy}{dx}$ in the other: thus

y	$\frac{dy}{dx}$
x^2	$2x$
x^3	$3x^2$
x^4	$4x^3$

Just look at these results: the operation of differentiating appears to have had the effect of diminishing the power of x by 1 (for example in the last case reducing x^4 to x^3), and at the same time multiplying by a number (the same number in fact which originally appeared as the power). Now, when you have once seen this, you might easily conjecture how the others will run. You would expect that differentiating x^5 would give $5x^4$, or differentiating x^6 would give $6x^5$. If you hesitate, try one of these, and see whether the conjecture comes right.

Try $y = x^5$.

$$\text{Then } y + dy = (x + dx)^5 = x^5 + 5x^4dx + 10x^3(dx)^2 + 10x^2(dx)^3 + 5x(dx)^4 + (dx)^5.$$

Neglecting all the terms containing small quantities of the higher orders, we have left

$$y + dy = x^5 + 5x^4dx$$

and subtracting $y = x^5$ leaves us

$$dy = 5x^4dx$$

whence $\dfrac{dy}{dx} = 5x^4$, exactly as we supposed.

Following out logically our observation, we should conclude that if we want to deal with any higher power—call it x^n—we could tackle it in the same way.

Let $$y = x^n$$

then we should expect to find that

$$\frac{dy}{dx} = nx^{n-1}$$

For example, let $n = 8$, then $y = x^8$; and differentiating it would give $\frac{dy}{dx} = 8x^7$.

And, indeed, the rule that differentiating x^n gives as the result nx^{n-1} is true for all cases where n is a whole number and positive. [Expanding $(x + dx)^n$ by the binomial theorem will at once show this.] But the question whether it is true for cases where n has negative or fractional values requires further consideration.

Case of a Negative Exponent.
Let $y = x^{-2}$. Then proceed as before:

$$y + dy = (x + dx)^{-2}$$
$$= x^{-2}\left(1 + \frac{dx}{x}\right)^{-2}$$

Expanding this by the binomial theorem, we get

$$= x^{-2}\left[1 - \frac{2dx}{x} + \frac{2(2+1)}{1 \times 2}\left(\frac{dx}{x}\right)^2 - \ldots\right]$$
$$= x^{-2} - 2x^{-3} \cdot dx + 3x^{-4}(dx)^2 - 4x^{-5}(dx)^3 + \text{etc.}$$

So, neglecting the small quantities of higher orders of smallness, we have:

$$y + dy = x^{-2} - 2x^{-3} \cdot dx$$

Subtracting the original $y = x^{-2}$, we find

$$dy = -2x^{-3}dx$$

$$\frac{dy}{dx} = -2x^{-3}$$

And this is still in accordance with the rule inferred above.

Case of a Fractional Exponent.
Let $y = x^{\frac{1}{2}}$. Then, as before,

$$y + dy = (x + dx)^{\frac{1}{2}} = x^{\frac{1}{2}}\left(1 + \frac{dx}{x}\right)^{\frac{1}{2}} = \sqrt{x}\left(1 + \frac{dx}{x}\right)^{\frac{1}{2}}$$

$$= \sqrt{x} + \frac{1}{2}\frac{dx}{\sqrt{x}} - \frac{1}{8}\frac{(dx)^2}{x\sqrt{x}} + \text{terms with higher powers of } dx.$$

Subtracting the original $y = x^{\frac{1}{2}}$, and neglecting higher powers we have left:

$$dy = \frac{1}{2}\frac{dx}{\sqrt{x}} = \frac{1}{2}x^{-\frac{1}{2}} \cdot dx$$

and $\qquad \dfrac{dy}{dx} = \dfrac{1}{2}x^{-\frac{1}{2}}$. This agrees with the general rule.

Summary. Let us see how far we have got. We have arrived at the following rule: To differentiate x^n, multiply it by the exponent and reduce the exponent by one, so giving us nx^{n-1} as the result.[3]

3. This rule is today known as the "power rule." It is the rule most frequently used in differentiating low-order functions.—M.G.

EXERCISES I

Differentiate the following:

(1) $y = x^{13}$

(2) $y = x^{-\frac{3}{2}}$

(3) $y = x^{2a}$

(4) $u = t^{2.4}$

(5) $z = \sqrt[3]{u}$

(6) $y = \sqrt[3]{x^{-5}}$

(7) $u = \sqrt[5]{\dfrac{1}{x^8}}$

(8) $y = 2x^a$

(9) $y = \sqrt[q]{x^3}$

(10) $y = \sqrt[n]{\dfrac{1}{x^m}}$

You have now learned how to differentiate powers of x. How easy it is!

Chapter V

NEXT STAGE. WHAT TO DO WITH CONSTANTS

In our equations we have regarded x as growing, and as a result of x being made to grow y also changed its value and grew. We usually think of x as a quantity that we can vary; and, regarding the variation of x as a sort of *cause*, we consider the resulting variation of y as an *effect*. In other words, we regard the value of y as depending on that of x. Both x and y are variables, but x is the one that we operate upon, and y is the "dependent variable". In all the preceding chapters we have been trying to find out rules for the proportion which the dependent variation in y bears to the variation independently made in x.

Our next step is to find out what effect on the process of differentiating is caused by the presence of *constants*, that is, of numbers which don't change when x or y changes its value.

Added Constants.

Let us begin with a simple case of an added constant, thus:

Let $$y = x^3 + 5$$

Just as before, let us suppose x to grow to $x + dx$ and y to grow to $y + dy$.

Then: $$\begin{aligned} y + dy &= (x + dx)^3 + 5 \\ &= x^3 + 3x^2dx + 3x(dx)^2 + (dx)^3 + 5 \end{aligned}$$

Neglecting the small quantities of higher orders, this becomes

$$y + dy = x^3 + 3x^2 \cdot dx + 5$$

Subtract the original $y = x^3 + 5$, and we have left:

$$dy = 3x^2 dx$$

$$\frac{dy}{dx} = 3x^2$$

So the 5 has quite disappeared. It added nothing to the growth of x, and does not enter into the derivative. If we had put 7, or 700, or any other number, instead of 5, it would have disappeared. So if we take the letter a, or b, or c to represent any constant, it will simply disappear when we differentiate.

If the additional constant had been of negative value, such as -5 or $-b$, it would equally have disappeared.

Multiplied Constants.

Take as a simple experiment this case:

Let $y = 7x^2$

Then on proceeding as before we get:

$$\begin{aligned} y + dy &= 7(x + dx)^2 \\ &= 7\{x^2 + 2x \cdot dx + (dx)^2\} \\ &= 7x^2 + 14x \cdot dx + 7(dx)^2 \end{aligned}$$

Then, subtracting the original $y = 7x^2$, and neglecting the last term, we have

$$dy = 14x \cdot dx$$

$$\frac{dy}{dx} = 14x$$

Let us illustrate this example by working out the graphs of the equations $y = 7x^2$ and $\dfrac{dy}{dx} = 14x$, by assigning to x a set of successive values, 0, 1, 2, 3, etc., and finding the corresponding values of y and of $\dfrac{dy}{dx}$.

These values we tabulate as follows:

x	0	1	2	3	4	5	−1	−2	−3
y	0	7	28	63	112	175	7	28	63
$\frac{dy}{dx}$	0	14	28	42	56	70	−14	−28	−42

Now plot these values to some convenient scale, and we obtain the two curves, Figs. 6 and 6*a*.

Carefully compare the two figures, and verify by inspection that the height of the ordinate of the derived curve, Fig. 6*a*, is porportional to the *slope* of the original curve, Fig. 6, at the corresponding value of x. To the left of the origin, where the original curve slopes negatively (that is, downward from left to right) the corresponding ordinates of the derived curve are negative.

Now, if we look back at previous pages, we shall see that simply differentiating x^2 gives us $2x$. So that the derivative of $7x^2$ is just 7 times as big as that of x^2. If we had taken $8x^2$, the derivative would have come out eight times as great as that of x^2. If we put $y = ax^2$, we shall get

$$\frac{dy}{dx} = a \times 2x$$

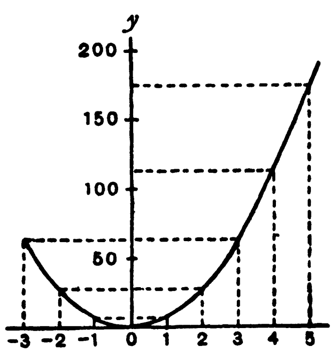

FIG. 6. Graph of $y = 7x^2$.

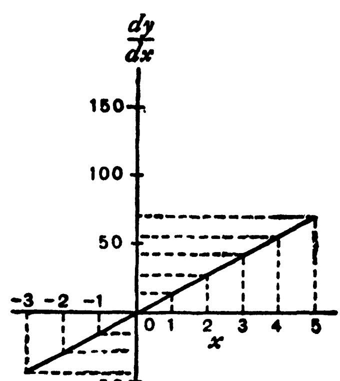

FIG. 6a. Graph of $\frac{dy}{dx} = 14x$.

If we had begun with $y = ax^n$, we should have had

$$\frac{dy}{dx} = a \times nx^{n-1}$$

So that any mere multiplication by a constant reappears as a mere multiplication when the thing is differentiated. And what is true about multiplication is equally true about *division:* for if, in the example above, we had taken as the constant $\frac{1}{7}$ instead of 7, we should have had the same $\frac{1}{7}$ come out in the result after differentiation.

Some Further Examples.

The following further examples, fully worked out, will enable you to master completely the process of differentiation as applied to ordinary algebraical expressions, and enable you to work out by yourself the examples given at the end of this chapter.

(1) Differentiate $y = \frac{x^5}{7} - \frac{3}{5}$

$-\frac{3}{5}$ is an added constant and vanishes.

We may then write at once

$$\frac{dy}{dx} = \frac{1}{7} \times 5 \times x^{5-1}$$

or

$$\frac{dy}{dx} = \frac{5}{7}\, x^4$$

(2) Differentiate $y = a\sqrt{x} - \frac{1}{2}\sqrt{a}$

The term $-\frac{1}{2}\sqrt{a}$ vanishes, being an added constant; and as $a\sqrt{x}$, in the index form, is written $ax^{\frac{1}{2}}$, we have

$$\frac{dy}{dx} = a \times \frac{1}{2} \times x^{\frac{1}{2}-1} = \frac{a}{2} \times x^{-\frac{1}{2}}$$

or

$$\frac{dy}{dx} = \frac{a}{2\sqrt{x}}$$

(3) The volume of a right circular cylinder of radius r and height h is given by the formula $V = \pi r^2 h$. Find the rate of variation of volume with the radius when $r = 5.5$ in. and $h = 20$ in. If $r = h$, find the dimensions of the cylinder so that a change of 1 in. in radius causes a change of 400 cubic inches in the volume.

The rate of variation of V with regard to r is

$$\frac{dV}{dr} = 2\pi rh$$

If $r = 5.5$ in. and $h = 20$ in. this becomes 691.2. It means that a change of radius of 1 inch will cause a change of volume of 691.2 cubic inches. This can be easily verified, for the volumes with $r = 5$ and $r = 6$ are 1570.8 cubic inches and 2262 cubic inches respectively, and $2262 - 1570.8 = 691.2$.

Also, if $h = r$, and h remains constant,

$$\frac{dV}{dr} = 2\pi r^2 = 400 \quad \text{and} \quad r = h = \sqrt{\frac{400}{2\pi}} = 7.98 \text{ in.}$$

If, however, $h = r$ and varies with r, then

$$\frac{dV}{dr} = 3\pi r^2 = 400 \quad \text{and} \quad r = h = \sqrt{\frac{400}{3\pi}} = 6.51 \text{ in.}$$

(4) The reading θ of a Féry's Radiation pyrometer is related to the centigrade temperature t of the observed body by the relation $\theta/\theta_1 = (t/t_1)^4$, where θ_1 is the reading corresponding to a known temperature t_1 of the observed body.

Compare the sensitivity of the pyrometer at temperatures 800° C., 1000° C., 1200° C., given that it read 25 when the temperature was 1000° C.

The sensitivity is the rate of variation of the reading with the temperature, that is, $\frac{d\theta}{dt}$. The formula may be written

$$\theta = \frac{\theta_1}{t_1{}^4}\, t^4 = \frac{25t^4}{1000^4}$$

and we have $$\frac{d\theta}{dt} = \frac{100t^3}{1000^4} = \frac{t^3}{10{,}000{,}000{,}000}$$

When $t = 800$, 1000 and 1200, we get $\dfrac{d\theta}{dt} = 0.0512$, 0.1 and 0.1728 respectively.

The sensitivity is approximately doubled from 800° to 1000°, and becomes three-quarters as great again up to 1200°.

EXERCISES II

Differentiate the following:

(1) $y = ax^3 + 6$ (2) $y = 13x^{\frac{3}{2}} - c$

(3) $y = 12x^{\frac{1}{2}} + c^{\frac{1}{2}}$ (4) $y = c^{\frac{1}{2}}x^{\frac{1}{2}}$

(5) $u = \dfrac{az^n - 1}{c}$ (6) $y = 1.18t^2 + 22.4$

Make up some other examples for yourself, and try your hand at differentiating them.

(7) If l_t and l_0 be the lengths of a rod of iron at the temperatures $t°$ C. and 0° C. respectively, then $l_t = l_0\,(1 + 0.000012t)$. Find the change of length of the rod per degree centigrade.

(8) It has been found that if c be the candle power of an incandescent electric lamp, and V be the voltage, $c = aV^b$, where a and b are constants.

Find the rate of change of the candle power with the voltage, and calculate the change of candle power per volt at 80, 100 and 120 volts in the case of a lamp for which $a = 0.5 \times 10^{-10}$ and $b = 6$.

(9) The frequency n of vibration of a string of diameter D, length L and specific gravity σ, stretched with a force T, is given by

$$n = \frac{1}{DL}\sqrt{\frac{gT}{\pi\sigma}}$$

Find the rate of change of the frequency when D, L, σ and T are varied singly.

(10) The greatest external pressure P which a tube can support without collapsing is given by

$$P = \left(\frac{2E}{1-\sigma^2}\right)\frac{t^3}{D^3}$$

where E and σ are constants, t is the thickness of the tube and D is its diameter. (This formula assumes that $4t$ is small compared to D.)

Compare the rate at which P varies for a small change of thickness and for a small change of diameter taking place separately.

(11) Find, from first principles, the rate at which the following vary with respect to a change in radius:

(a) the circumference of a circle of radius r;
(b) the area of a circle of radius r;
(c) the lateral area of a cone of slant dimension l;
(d) the volume of a cone of radius r and height h;
(e) the area of a sphere of radius r;
(f) the volume of a sphere of radius r.

Chapter VI

SUMS, DIFFERENCES, PRODUCTS, AND QUOTIENTS[1]

We have learned how to differentiate simple algebraical functions such as $x^2 + c$ or ax^4, and we have now to consider how to tackle the *sum* of two or more functions.

For instance, let

$$y = (x^2 + c) + (ax^4 + b)$$

what will its $\frac{dy}{dx}$ be? How are we to go to work on this new job?

The answer to this question is quite simple: just differentiate them, one after the other, thus:

$$\frac{dy}{dx} = 2x + 4ax^3$$

If you have any doubt whether this is right, try a more general case, working it by first principles. And this is the way.

Let $y = u + v$, where u is any function of x, and v any other function of x. Then, letting x increase to $x + dx$, y will increase to $y + dy$; and u will increase to $u + du$; and v to $v + dv$.

And we shall have:

$$y + dy = u + du + v + dv$$

1. The rules given in this chapter are today known as the sum rule, the difference rule, the product rule, and the quotient rule.—M.G.

Subtracting the original $y = u + v$, we get

$$dy = du + dv$$

and dividing through by dx, we get:

$$\frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}$$

This justifies the procedure. You differentiate each function separately and add the results. So if now we take the example of the preceding paragraph, and put in the values of the two functions, we shall have, using the notation shown.

$$\begin{aligned} \frac{dy}{dx} &= \frac{d(x^2 + c)}{dx} + \frac{d(ax^4 + b)}{dx} \\ &= 2x \qquad\quad + 4ax^3 \end{aligned}$$

exactly as before.

If there were three functions of x, which we may call u, v and w, so that

$$y = u + v + w$$

then $$\frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx} + \frac{dw}{dx}$$

As for the rule about *subtraction,* it follows at once; for if the function v had itself had a negative sign, its derivative would also be negative; so that by differentiating

$$y = u - v$$

we should get $$\frac{dy}{dx} = \frac{du}{dx} - \frac{dv}{dx}$$

But when we come to do with *Products,* the thing is not quite so simple.

Suppose we were asked to differentiate the expression

$$y = (x^2 + c) \times (ax^4 + b)$$

what are we to do? The result will certainly *not* be $2x \times 4ax^3$; for it is easy to see that neither $c \times ax^4$, nor $x^2 \times b$, would have been taken into that product.

Now there are two ways in which we may go to work.

First Way. Do the multiplying first, and, having worked it out, then differentiate.

Accordingly, we multiply together $x^2 + c$ and $ax^4 + b$.

This gives $ax^6 + acx^4 + bx^2 + bc$.

Now differentiate, and we get:

$$\frac{dy}{dx} = 6ax^5 + 4acx^3 + 2bx$$

Second Way. Go back to first principles, and consider the equation

$$y = u \times v$$

where u is one function of x, and v is any other function of x. Then, if x grows to be $x + dx$; and y to $y + dy$; and u becomes $u + du$; and v becomes $v + dv$, we shall have:

$$\begin{aligned} y + dy &= (u + du) \times (v + dv) \\ &= u \cdot v + u \cdot dv + v \cdot du + du \cdot dv \end{aligned}$$

Now $du \cdot dv$ is a small quantity of the second order of smallness, and therefore in the limit may be discarded, leaving

$$y + dy = u \cdot v + u \cdot dv + v \cdot du$$

Then, subtracting the original $y = u \cdot v$, we have left

$$dy = u \cdot dv + v \cdot du$$

and, dividing through by dx, we get the result:

$$\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$$

This shows that our instructions will be as follows: *To differentiate the product of two functions, multiply each function by the derivative of the other, and add together the two products so obtained.*

You should note that this process amounts to the following:

Treat u as constant while you differentiate v; then treat v as constant while you differentiate u; and the whole derivative $\frac{dy}{dx}$ will be the sum of the results of these two treatments.

Now, having found this rule, apply it to the concrete example which was considered above.

We want to differentiate the product

$$(x^2 + c) \times (ax^4 + b)$$

Call $\qquad (x^2 + c) = u; \quad \text{and} \quad (ax^4 + b) = v$

Then, by the general rule just established, we may write:

$$\begin{aligned}
\frac{dy}{dx} &= (x^2 + c)\frac{d(ax^4 + b)}{dx} + (ax^4 + b)\frac{d(x^2 + c)}{dx} \\
&= (x^2 + c)4ax^3 \qquad + (ax^4 + b)2x \\
&= 4ax^5 + 4acx^3 \qquad + 2ax^5 + 2bx \\
\frac{dy}{dx} &= 6ax^5 + 4acx^3 \qquad + 2bx
\end{aligned}$$

Exactly as before.

Lastly, we have to differentiate *quotients.*

Think of this example, $y = \dfrac{bx^5 + c}{x^2 + a}$. In such a case it is no use to try to work out the division beforehand, because $x^2 + a$ will not divide into $bx^5 + c$, neither have they any common factor. So there is nothing for it but to go back to first principles, and find a rule.

So we will put $\qquad y = \dfrac{u}{v}$

where u and v are two different functions of the independent variable x. Then, when x becomes $x + dx$, y will become $y + dy$; and u will become $u + du$; and v will become $v + dv$. So then

$$y + dy = \frac{u + du}{v + dv}$$

Now perform the algebraic division, thus:

$$
\begin{array}{r|l}
v+dv & u+du \qquad \left| \dfrac{u}{v}+\dfrac{du}{v}-\dfrac{u\cdot dv}{v^2} \right. \\
\hline
 & u+\dfrac{u\cdot dv}{v} \\
\cline{2-2}
 & \qquad du-\dfrac{u\cdot dv}{v} \\
 & \qquad du+\dfrac{du\cdot dv}{v} \\
\cline{2-2}
 & \qquad\quad -\dfrac{u\cdot dv}{v}-\dfrac{du\cdot dv}{v} \\
 & \qquad\quad -\dfrac{u\cdot dv}{v}-\dfrac{u\cdot dv\cdot dv}{v^2} \\
\cline{2-2}
 & \qquad\qquad\qquad -\dfrac{du\cdot dv}{v}+\dfrac{u\cdot dv\cdot dv}{v^2}
\end{array}
$$

As both these remainders are small quantities of the second order, they may be neglected, and the division may stop here, since any further remainders would be of still smaller magnitudes.

So we have got:

$$y+dy=\frac{u}{v}+\frac{du}{v}-\frac{u\cdot dv}{v^2}$$

which may be written

$$=\frac{u}{v}+\frac{v\cdot du-u\cdot dv}{v^2}$$

Now subtract the original $y=\frac{u}{v}$, and we have left:

$$dy=\frac{v\cdot du-u\cdot dv}{v^2}$$

whence

$$\frac{dy}{dx}=\frac{v\dfrac{du}{dx}-u\dfrac{dv}{dx}}{v^2}$$

This gives us our instructions as to *how to differentiate a quotient of* two functions. *Multiply the divisor function by the derivative of the dividend function; then multiply the dividend function by the derivative of the divisor function; and subtract the latter product from the former. Lastly, divide the difference by the square of the divisor function.*

Going back to our example $y = \dfrac{bx^5 + c}{x^2 + a}$

write $\qquad bx^5 + c = u;$ and $x^2 + a = v$

Then
$$\frac{dy}{dx} = \frac{(x^2 + a)\dfrac{d(bx^5 + c)}{dx} - (bx^5 + c)\dfrac{d(x^2 + a)}{dx}}{(x^2 + a)^2}$$

$$= \frac{(x^2 + a)(5bx^4) - (bx^5 + c)(2x)}{(x^2 + a)^2}$$

$$\frac{dy}{dx} = \frac{3bx^6 + 5abx^4 - 2cx}{(x^2 + a)^2}$$

The working out of quotients is often tedious, but there is nothing difficult about it.

Some further examples fully worked out are given hereafter.

(1) Differentiate $\quad y = \dfrac{a}{b^2}x^3 - \dfrac{a^2}{b}x + \dfrac{a^2}{b^2}$

Being a constant, $\dfrac{a^2}{b^2}$ vanishes, and we have

$$\frac{dy}{dx} = \frac{a}{b^2} \times 3 \times x^{3-1} - \frac{a^2}{b} \times 1 \times x^{1-1}$$

But $x^{1-1} = x^0 = 1$; so we get:

$$\frac{dy}{dx} = \frac{3a}{b^2}x^2 - \frac{a^2}{b}$$

(2) Differentiate $\quad y = 2a\sqrt{bx^3} - \dfrac{3b\sqrt[3]{a}}{x} - 2\sqrt{ab}$

Putting x in the exponent form, we get

$$y = 2a\sqrt{b}\,x^{\frac{3}{2}} - 3b\sqrt[3]{a}\,x^{-1} - 2\sqrt{ab}$$

Now $$\frac{dy}{dx} = 2a\sqrt{b} \times \tfrac{3}{2} \times x^{\frac{3}{2}-1} - 3b\sqrt[3]{a} \times (-1) \times x^{-1-1}$$

or, $$\frac{dy}{dx} = 3a\sqrt{bx} + \frac{3b\sqrt[3]{a}}{x^2}$$

(3) Differentiate $\quad z = 1.8\sqrt[3]{\dfrac{1}{\theta^2}} - \dfrac{4.4}{\sqrt[5]{\theta}} - 27.$

This may be written: $z = 1.8\theta^{-\frac{2}{3}} - 4.4\theta^{-\frac{1}{5}} - 27.$
The 27 vanishes, and we have

$$\frac{dz}{d\theta} = 1.8 \times \left(-\tfrac{2}{3}\right)\theta^{-\frac{2}{3}-1} - 4.4 \times \left(-\tfrac{1}{5}\right)\theta^{-\frac{1}{5}-1}$$

or, $$\frac{dz}{d\theta} = -1.2\theta^{-\frac{5}{3}} + 0.88\theta^{-\frac{6}{5}}$$

or, $$\frac{dz}{d\theta} = \frac{0.88}{\sqrt[5]{\theta^6}} - \frac{1.2}{\sqrt[3]{\theta^5}}$$

(4) Differentiate $\quad v = (3t^2 - 1.2t + 1)^3.$

A direct way of doing this will be explained later; but we can nevertheless manage it now without any difficulty.

Developing the cube, we get

$$v = 27t^6 - 32.4t^5 + 39.96t^4 - 23.328t^3 + 13.32t^2 - 3.6t + 1$$

hence

$$\frac{dv}{dt} = 162t^5 - 162t^4 + 159.84t^3 - 69.984t^2 + 26.64t - 3.6.$$

(5) Differentiate $\quad y = (2x - 3)(x + 1)^2.$

$$\frac{dy}{dx} = (2x-3)\frac{d[(x+1)(x+1)]}{dx} + (x+1)^2\frac{d(2x-3)}{dx}$$

$$= (2x-3)\left[(x+1)\frac{d(x+1)}{dx} + (x+1)\frac{d(x+1)}{dx}\right] + (x+1)^2\frac{d(2x-3)}{dx}$$

$$= 2(x+1)[(2x-3)+(x+1)] = 2(x+1)(3x-2)$$

or, more simply, multiply out and then differentiate.

(6) Differentiate $y = 0.5x^3(x-3)$.

$$\frac{dy}{dx} = 0.5\left[x^3\frac{d(x-3)}{dx} + (x-3)\frac{d(x^3)}{dx}\right]$$

$$= 0.5[x^3 + (x-3)\times 3x^2] = 2x^3 - 4.5x^2$$

Same remarks as for preceding example.

(7) Differentiate $w = \left(\theta + \dfrac{1}{\theta}\right)\left(\sqrt{\theta} + \dfrac{1}{\sqrt{\theta}}\right)$

This may be written

$$w = (\theta + \theta^{-1})(\theta^{\frac{1}{2}} + \theta^{-\frac{1}{2}})$$

$$\frac{dw}{d\theta} = (\theta + \theta^{-1})\frac{d(\theta^{\frac{1}{2}} + \theta^{-\frac{1}{2}})}{d\theta} + (\theta^{\frac{1}{2}} + \theta^{-\frac{1}{2}})\frac{d(\theta + \theta^{-1})}{d\theta}$$

$$= (\theta + \theta^{-1})(\tfrac{1}{2}\theta^{-\frac{1}{2}} - \tfrac{1}{2}\theta^{-\frac{3}{2}}) + (\theta^{\frac{1}{2}} + \theta^{-\frac{1}{2}})(1 - \theta^{-2})$$

$$= \tfrac{1}{2}(\theta^{\frac{1}{2}} + \theta^{-\frac{3}{2}} - \theta^{-\frac{1}{2}} - \theta^{-\frac{5}{2}}) + (\theta^{\frac{1}{2}} + \theta^{-\frac{1}{2}} - \theta^{-\frac{3}{2}} - \theta^{-\frac{5}{2}})$$

$$= \frac{3}{2}\left(\sqrt{\theta} - \frac{1}{\sqrt{\theta^5}}\right) + \frac{1}{2}\left(\frac{1}{\sqrt{\theta}} - \frac{1}{\sqrt{\theta^3}}\right)$$

This, again, could be obtained more simply by multiplying the two factors first, and differentiating afterwards. This is not, however, always possible; see, for instance, example 8 in Chapter XV, in which the rule for differentiating a product *must* be used.

(8) Differentiate $y = \dfrac{a}{1 + a\sqrt{x} + a^2x}$

$$\frac{dy}{dx} = \frac{\left(1 + ax^{\frac{1}{2}} + a^2x\right) \times 0 - a\dfrac{d\left(1 + ax^{\frac{1}{2}} + a^2x\right)}{dx}}{\left(1 + a\sqrt{x} + a^2x\right)^2}$$

$$= -\frac{a\left(\frac{1}{2}ax^{-\frac{1}{2}} + a^2\right)}{(1 + ax^{\frac{1}{2}} + a^2x)^2}$$

(9) Differentiate $y = \dfrac{x^2}{x^2 + 1}$.

$$\frac{dy}{dx} = \frac{(x^2 + 1)2x - x^2 \times 2x}{(x^2 + 1)^2} = \frac{2x}{(x^2 + 1)^2}$$

(10) Differentiate $y = \dfrac{a + \sqrt{x}}{a - \sqrt{x}}$

In the exponent form, $y = \dfrac{a + x^{\frac{1}{2}}}{a - x^{\frac{1}{2}}}$

$$\frac{dy}{dx} = \frac{\left(a - x^{\frac{1}{2}}\right)\left(\frac{1}{2}x^{-\frac{1}{2}}\right) - \left(a + x^{\frac{1}{2}}\right)\left(-\frac{1}{2}x^{-\frac{1}{2}}\right)}{\left(a - x^{\frac{1}{2}}\right)^2} = \frac{a - x^{\frac{1}{2}} + a + x^{\frac{1}{2}}}{2\left(a - x^{\frac{1}{2}}\right)^2 x^{\frac{1}{2}}}$$

hence $$\frac{dy}{dx} = \frac{a}{\left(a - \sqrt{x}\right)^2\sqrt{x}}$$

(11) Differentiate $\theta = \dfrac{1 - a\sqrt[3]{t^2}}{1 + a\sqrt[2]{t^3}}$

Now $$\theta = \frac{1 - at^{\frac{2}{3}}}{1 + at^{\frac{3}{2}}}$$

$$\frac{d\theta}{dt}=\frac{\left(1+at^{\frac{3}{2}}\right)\left(-\frac{2}{3}at^{-\frac{1}{3}}\right)-\left(1-at^{\frac{2}{3}}\right)\times\frac{3}{2}at^{\frac{1}{2}}}{\left(1+at^{\frac{3}{2}}\right)^2}$$

$$=\frac{5a^2\sqrt[6]{t^7}-\dfrac{4a}{\sqrt[3]{t}}-9a\sqrt[2]{t}}{6\left(1+a\sqrt[2]{t^3}\right)^2}$$

(12) A reservoir of square cross-section has sides sloping at an angle of 45° with the vertical. The side of the bottom is p feet in length, and water flows in the reservoir at the rate of c cubic feet per minute. Find an expression for the rate at which the surface of the water is rising at the instant its depth is h feet. Calculate this rate when $p=17$, $h=4$ and $c=35$.

The volume of a frustum of pyramid of height H, and of bases A and a, is $V=\dfrac{H}{3}\left(A+a+\sqrt{Aa}\right)$. It is easily seen that, the slope being 45°, for a depth of h, the length of the side of the upper square surface of water is $(p+2h)$ feet; thus $A=p^2$, $a=(p+2h)^2$ and the volume of the water is

$$\tfrac{1}{3}h\{p^2+p(p+2h)+(p+2h)^2\}\text{ cubic feet}$$
$$=p^2h+2ph^2+\tfrac{4}{3}h^3\text{ cubic feet}$$

Now, if t be the time in minutes taken for this volume of water to flow in,

$$ct=p^2h+2ph^2+\tfrac{4}{3}h^3$$

From this relation we have the rate at which h increases with t, that is $\dfrac{dh}{dt}$, but as the above expression is in the form, $t=$ function of h, rather than $h=$ function of t, it will be easier to find $\dfrac{dt}{dh}$ and then invert the result, for

$$\frac{dt}{dh}\times\frac{dh}{dt}=1$$

Hence, since c and p are constants, and

$$ct = p^2 h + 2ph^2 + \tfrac{4}{3}h^3$$

$$c\frac{dt}{dh} = p^2 + 4ph + 4h^2 = (p + 2h)^2$$

so that $\dfrac{dh}{dt} = \dfrac{c}{(p + 2h)^2}$, which is the required expression.

When $p = 17$, $h = 4$ and $c = 35$; this becomes 0.056 feet per minute.

(13) The absolute pressure, in atmospheres, P, of saturated steam at the temperature t° C. is $P = \left(\dfrac{40 + t}{140}\right)^5$ as long as t is above 80°. Find the rate of variation of the pressure with the temperature at 100° C.

Since $$P = \left(\frac{40 + t}{140}\right)^5; \quad \frac{dP}{dt} = \frac{5(40 + t)^4}{(140)^5}$$

so that when $t = 100$,

$$\frac{dP}{dt} = \frac{5 \times (140)^4}{(140)^5} = \frac{5}{140} = \frac{1}{28} = 0.036$$

Thus, the rate of variation of the pressure is, when $t = 100$, 0.036 atmosphere per degree centigrade change of temperature.

EXERCISES III

(1) Differentiate

(*a*) $u = 1 + x + \dfrac{x^2}{1 \times 2} + \dfrac{x^3}{1 \times 2 \times 3} + \dots$

(*b*) $y = ax^2 + bx + c$ (*c*) $y = (x + a)^2$

(*d*) $y = (x + a)^3$

(2) If $w = at - \frac{1}{2}bt^2$, find $\dfrac{dw}{dt}$

(3) Find the derivative of

$$y = \left(x + \sqrt{-1}\right) \times \left(x - \sqrt{-1}\right)$$

(4) Differentiate

$$y = (197x - 34x^2) \times (7 + 22x - 83x^3)$$

(5) If $x = (y + 3) \times (y + 5)$, find $\dfrac{dx}{dy}$.

(6) Differentiate $y = 1.3709x \times (112.6 + 45.202x^2)$.
Find the derivatives of

(7) $y = \dfrac{2x + 3}{3x + 2}$ (8) $y = \dfrac{1 + x + 2x^2 + 3x^3}{1 + x + 2x^2}$

(9) $y = \dfrac{ax + b}{cx + d}$ (10) $y = \dfrac{x^n + a}{x^{-n} + b}$

(11) The temperature t of the filament of an incandescent electric lamp is connected to the current passing through the lamp by the relation

$$C = a + bt + ct^2$$

Find an expression giving the variation of the current corresponding to a variation of temperature.

(12) The following formulae have been proposed to express the relation between the electric resistance R of a wire at the temperature $t°$ C., and the resistance R_0 of that same wire at 0° centigrade, a and b being constants.

$$R = R_0(1 + at + bt^2)$$

$$R = R_0\left(1 + at + b\sqrt{t}\right)$$

$$R = R_0(1 + at + bt^2)^{-1}$$

Find the rate of variation of the resistance with regard to temperature as given by each of these formulae.

(13) The electromotive force E of a certain type of standard cell has been found to vary with the temperature t according to the relation

$$E = 1.4340[1 - 0.000814(t - 15) \\ + 0.000007(t - 15)^2] \text{ volts}$$

Find the change of electromotive force per degree, at 15°, 20° and 25°.

(14) The electromotive force necessary to maintain an electric arc of length l with a current intensity i has been found to be

$$E = a + bl + \frac{c + kl}{i}$$

where a, b, c, k are constants.

Find an expression for the variation of the electromotive force (*a*) with regard to the length of the arc; (*b*) with regard to the strength of the current.

Chapter VII

SUCCESSIVE DIFFERENTIATION

Let us try the effect of repeating several times over the operation of differentiating a function. Begin with a concrete case.

Let $y = x^5$.

First differentiation,	$5x^4$.	
Second differentiation,	$5 \times 4x^3$	$= 20x^3$.
Third differentiation,	$5 \times 4 \times 3x^2$	$= 60x^2$.
Fourth differentiation,	$5 \times 4 \times 3 \times 2x$	$= 120x$.
Fifth differentiation,	$5 \times 4 \times 3 \times 2 \times 1$	$= 120$.
Sixth differentiation,		$= 0$.[1]

1. When applied to an object moving at a constant rate, the first derivative gives its change of position per second. If the object is accelerating, the second derivative gives the rate at which the first derivative is changing, that is, its change of position per second per second. If the acceleration is changing, the third derivative gives the rate at which the second derivative is changing, namely the change of position per second per second per second. Physicists call such a change a "jerk," such as the way an old car jerks if there is too sudden a change in how it is accelerating.

Second derivatives, in which time is the independent variable, turn up everywhere in physics, less often in other sciences. In economics, a second derivative can express the rate at which a worker's annual increase in wages is increasing (or decreasing). Third derivatives are also useful in many branches of physics. Beyond the third, higher order derivatives are seldom needed. This testifies to the fortunate fact that the universe seems to favor simplicity in its fundamental laws.—M.G.

There is a certain notation, with which we are already acquainted, used by some writers, that is very convenient. This is to employ the general symbol $f(x)$ for any function of x. Here the symbol $f(\ )$ is read as "function of", without saying what particular function is meant. So the statement $y = f(x)$ merely tells us that y is a function of x, it may be x^2 or ax^n, or $\cos x$ or any other complicated function of x.

The corresponding symbol for the derivative is $f'(x)$, which is simpler to write than $\dfrac{dy}{dx}$. This is called the "derived function" of x.

Suppose we differentiate over again, we shall get the "second derived function" or second derivative which is denoted by $f''(x)$; and so on.

Now let us generalize.

Let $y = f(x) = x^n$.

First differentiation, $f'(x) = nx^{n-1}$.
Second differentiation, $f''(x) = n(n-1)x^{n-2}$.
Third differentiation, $f'''(x) = n(n-1)(n-2)x^{n-3}$.
Fourth differentiation, $f''''(x) = n(n-1)(n-2)(n-3)x^{n-4}$.
etc., etc.

But this is not the only way of indicating successive differentiations. For, if the original function be

$$y = f(x);$$

differentiating once gives $$\frac{dy}{dx} = f'(x);$$

differentiating twice gives $$\frac{d\left(\dfrac{dy}{dx}\right)}{dx} = f''(x);$$

and this is more conveniently written as $\dfrac{d^2y}{(dx)^2}$, or more usually

$\frac{d^2y}{dx^2}$. Similarly, we may write as the result of differentiating three times, $\frac{d^3y}{dx^3} = f'''(x)$.

Examples.

Now let us try $y = f(x) = 7x^4 + 3.5x^3 - \frac{1}{2}x^2 + x - 2$

$$\frac{dy}{dx} = f'(x) \quad = 28x^3 + 10.5x^2 - x + 1$$

$$\frac{d^2y}{dx^2} = f''(x) \quad = 84x^2 + 21x - 1$$

$$\frac{d^3y}{dx^3} = f'''(x) \quad = 168x + 21$$

$$\frac{d^4y}{dx^4} = f''''(x) \quad = 168$$

$$\frac{d^5y}{dx^5} = f'''''(x) = 0$$

In a similar manner if $y = \phi(x) = 3x(x^2 - 4)$

$$\phi'(x) = \frac{dy}{dx} \quad = 3[x \times 2x + (x^2 - 4) \times 1] = 3(3x^2 - 4)$$

$$\phi''(x) = \frac{d^2y}{dx^2} = 3 \times 6x = 18x$$

$$\phi'''(x) = \frac{d^3y}{dx^3} = 18$$

$$\phi''''(x) = \frac{d^4y}{dx^4} = 0$$

EXERCISES IV

Find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for the following expressions:

(1) $y = 17x + 12x^2$ (2) $y = \frac{x^2 + a}{x + a}$

(3) $y = 1 + \frac{x}{1} + \frac{x^2}{1 \times 2} + \frac{x^3}{1 \times 2 \times 3} + \frac{x^4}{1 \times 2 \times 3 \times 4}$

(4) Find the 2nd and 3rd derivatives in Exercises III, No. 1 to No. 7, and in the Examples given in Chapter VI, No. 1 to No. 7.

Chapter VIII

WHEN TIME VARIES

Some of the most important problems of the calculus are those where time is the independent variable, and we have to think about the values of some other quantity that varies when the time varies. Some things grow larger as time goes on; some other things grow smaller. The distance that a train has travelled from its starting place goes on ever increasing as time goes on. Trees grow taller as the years go by. Which is growing at the greater rate: a plant 12 inches high which in one month becomes 14 inches high, or a tree 12 feet high which in a year becomes 14 feet high?

In this chapter we are going to make much use of the word *rate*. Nothing to do with birth rate or death rate, though these words suggest so many births or deaths per thousand of the population. When a car whizzes by us, we say: What a terrific rate! When a spendthrift is flinging about his money, we remark that that young man is living at a prodigious rate. What do we mean by *rate*? In both these cases we are making a mental comparison of something that is happening, and the length of time it takes to happen. If the car goes 10 yards per second, a simple bit of mental arithmetic will show us that this is equivalent—while it lasts—to a rate of 600 yards per minute, or over 20 miles per hour.

Now in what sense is it true that a speed of 10 yards per second is the same as 600 yards per minute? Ten yards is not the same as 600 yards, nor is one second the same thing as one minute. What we mean by saying that the *rate* is the same, is this: that the proportion borne between distance passed over and time taken to pass over it, is the same in both cases.

Now try to put some of these ideas into differential notation. Let y in this case stand for money, and let t stand for time.

If you are spending money, and the amount you spend in a short time dt be called dy, the *rate* of spending it will be $\frac{dy}{dt}$; or, as regards saving, with a minus sign, as $-\frac{dy}{dt}$, because then dy is a *decrement*, not an increment. But money is not a good example for the calculus, because it generally comes and goes by jumps, not by a continuous flow—you may earn $20,000 a year, but it does not keep running in all day long in a thin stream; it comes in only weekly, or monthly, or quarterly, in lumps: and your expenditure also goes out in sudden payments.

A more apt illustration of the idea of a rate is furnished by the speed of a moving body. From London to Liverpool is 200 miles. If a train leaves London at 7 o'clock, and reaches Liverpool at 11 o'clock, you know that, since it has travelled 200 miles in 4 hours, its average rate must have been 50 miles per hour; because $\frac{200}{4} = \frac{50}{1}$. Here you are really making a mental comparison between the distance passed over and the time taken to pass over it. You are dividing one by the other. If y is the whole distance, and t the whole time, clearly the average rate is $\frac{y}{t}$. Now the speed was not actually constant all the way: at starting, and during the slowing up at the end of the journey, the speed was less. Probably at some part, when running downhill, the speed was over 60 miles an hour. If, during any particular element of time dt, the corresponding element of distance passed over was dy, then at that part of the journey the speed was $\frac{dy}{dt}$. The *rate* at which one quantity (in the present instance, *distance*) is changing in relation to the other quantity (in this case, *time*) is properly expressed, then, by stating the derivative of one with respect to the other. A *velocity*, scientifically expressed, is the rate at which a very small distance in any given direction is being passed over, and may therefore be written

$$v = \frac{dy}{dt}$$

But if the velocity v is not uniform, then it must be either increasing or else decreasing. The rate at which a velocity is increasing is called the *acceleration.* If a moving body is, at any particular instant, gaining an additional velocity dv in an element of time dt, then the acceleration a at that instant may be written

$$a = \frac{dv}{dt}$$

But since $v = \dfrac{dy}{dt}$,

$$a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dy}{dt}\right)$$

which is usually written $a = \dfrac{d^2y}{dt^2}$;

or the acceleration is the second derivative of the distance, with respect to time. Acceleration is expressed as a change of velocity in unit time, for instance, as being so many feet per second per second; the notation used being ft/sec^2.

When a railway train has just begun to move, its velocity v is small; but it is rapidly gaining speed—it is being hurried up, or accelerated, by the effort of the engine. So its $\dfrac{d^2y}{dt^2}$ is large. When it has got up its top speed it is no longer being accelerated, so that then $\dfrac{d^2y}{dt^2}$ has fallen to zero. But when it nears its stopping place its speed begins to slow down; may, indeed, slow down very quickly if the brakes are put on, and during this period of *deceleration* or slackening of pace, the value of $\dfrac{dv}{dt}$, that is, of $\dfrac{d^2y}{dt^2}$, will be negative.

To accelerate a mass m requires the continuous application of force. The force necessary to accelerate a mass is proportional to the mass, and it is also proportional to the acceleration which is being imparted. Hence we may write for the force f, the expression

$$f = ma$$

or $$f = m\frac{dv}{dt}$$

or $$f = m\frac{d^2y}{dt^2}$$

The product of a mass by the speed at which it is going is called its *momentum*, and is in symbols mv. If we differentiate momentum with respect to time we shall get $\frac{d(mv)}{dt}$ for the rate of change of momentum. But, since m is a constant quantity, this may be written $m\,\frac{dv}{dt}$, which we see above is the same as f. That is to say, force may be expressed either as mass times acceleration, or as rate of change of momentum.

Again, if a force is employed to move something (against an equal and opposite counter-force), it does *work*; and the amount of work done is measured by the product of the force into the distance (in its own direction) through which its point of application moves forward. So if a force f moves forward through a length y, the work done (which we may call w) will be

$$w = f \times y$$

where we take f as a constant force. If the force varies at different parts of the range y, then we must find an expression for its value from point to point. If f be the force along the small element of length dy, the amount of work done will be $f \times dy$. But as dy is only an element of length, only an element of work will be done. If we write w for work, then an element of work will be dw; and we have

$$dw = f \times dy$$

which may be written $$dw = ma \cdot dy;$$

or $$dw = m\frac{d^2y}{dt^2} \cdot dy$$

or $$dw = m\frac{dv}{dt} \cdot dy$$

Further, we may transpose the expression and write

$$\frac{dw}{dy} = f$$

This gives us yet a third definition of *force;* that if it is being used to produce a displacement in any direction, the force (in that direction) is equal to the rate at which work is being done per unit of length in that direction. In this last sentence the word *rate* is clearly not used in its time-sense, but in its meaning as ratio or proportion.

Sir Isaac Newton, who was (along with Leibniz) an inventor of the methods of the calculus, regarded all quantities that were varying as *flowing;* and the ratio which we nowadays call the derivative he regarded as the rate of flowing, or the *fluxion* of the quantity in question. He did not use the notation of the dy and dx, and dt (this was due to Leibniz), but had instead a notation of his own. If y was a quantity that varied, or "flowed", then his symbol for its rate of variation (or "fluxion") was $\dot{y}$. If x was the variable, then its fluxion was called $\dot{x}$. The dot over the letter indicated that it had been differentiated. But this notation does not tell us what is the independent variable with respect to which the differentiation has been effected. When we see $\frac{dy}{dt}$ we know that y is to be differentiated with respect to t. If we see $\frac{dy}{dx}$ we know that y is to be differentiated with respect to x. But if we see merely $\dot{y}$, we cannot tell without looking at the context whether this is to mean $\frac{dy}{dx}$ or $\frac{dy}{dt}$ or $\frac{dy}{dz}$, or what is the other variable. So, therefore, this fluxional notation is less informing than the differential notation, and has in consequence largely dropped out of use. But its simplicity gives it an advantage if only we will agree to use it for those cases exclusively where *time* is the independent variable. In that case $\dot{y}$ will mean $\frac{dy}{dt}$ and $\dot{u}$ will mean $\frac{du}{dt}$; and $\ddot{x}$ will mean $\frac{d^2x}{dt^2}$.

Adopting this fluxional notation we may write the mechanical equations considered in the paragraphs above, as follows:

distance	x
velocity	$v = \dot{x}$
acceleration	$a = \dot{v} = \ddot{x}$
force	$f = m\dot{v} = m\ddot{x}$
work	$w = x \times m\ddot{x}$

Examples.

(1) A body moves so that the distance x (in feet), which it travels from a certain point O, is given by the relation

$$x = 0.2t^2 + 10.4$$

where t is the time in seconds elapsed since a certain instant. Find the velocity and acceleration 5 seconds after the body began to move, and also find the corresponding values when the distance covered is 100 feet. Find also the average velocity during the first 10 seconds of its motion. (Suppose distances and motion to the right to be positive.)

Now $$x = 0.2t^2 + 10.4$$

$$v = \dot{x} = \frac{dx}{dt} = 0.4t; \quad \text{and} \quad a = \ddot{x} = \frac{d^2x}{dt^2} = 0.4 = \text{constant.}$$

When $t = 0$, $x = 10.4$ and $v = 0$. The body started from a point 10.4 feet to the right of the point O; and the time was reckoned from the instant the body started.

When $t = 5$, $v = 0.4 \times 5 = 2$ ft./sec.; $a = 0.4$ ft./sec^2.

When $x = 100$, $100 = 0.2t^2 + 10.4$, or $t^2 = 448$,

and $t = 21.17$ sec.; $v = 0.4 \times 21.17 = 8.468$ ft./sec.

When $t = 10$,

distance travelled $= 0.2 \times 10^2 + 10.4 - 10.4 = 20$ ft.

Average velocity $= \frac{20}{10} = 2$ ft./sec.

(It is the same velocity as the velocity at the middle of the interval, $t = 5$; for, the acceleration being constant, the velocity has varied uniformly from zero when $t = 0$ to 4 ft./sec. when $t = 10$.)

(2) In the above problem let us suppose

$$x = 0.2t^2 + 3t + 10.4$$

$$v = \dot{x} = \frac{dx}{dt} = 0.4t + 3; \quad a = \ddot{x} = \frac{d^2x}{dt^2} = 0.4 = \text{constant.}$$

When $t = 0$, $x = 10.4$ and $v = 3$ ft./sec., the time is reckoned from the instant at which the body passed a point 10.4 ft. from the point O, its velocity being then already 3 ft./sec. To find the time elapsed since it began moving, let $v = 0$; then $0.4t + 3 = 0$, $t = -\frac{3}{.4} = -7.5$ sec. The body began moving 7.5 sec. before time was begun to be observed; 5 seconds after this gives $t = -2.5$ and $v = 0.4 \times -2.5 + 3 = 2$ ft./sec.

When $x = 100$ ft.,

$$100 = 0.2t^2 + 3t + 10.4; \quad \text{or} \quad t^2 + 15t - 448 = 0$$

hence $\quad t = 14.96$ sec., $v = 0.4 \times 14.96 + 3 = 8.98$ ft./sec.

To find the distance travelled during the first 10 seconds of the motion one must know how far the body was from the point O when it started.

When $t = -7.5$,

$$x = 0.2 \times (-7.5)^2 - 3 \times 7.5 + 10.4 = -0.85 \text{ ft.}$$

that is 0.85 ft. to the left of the point O.

Now, when $t = 2.5$,

$$x = 0.2 \times 2.5^2 + 3 \times 2.5 + 10.4 = 19.15$$

So, in 10 seconds, the distance travelled was $19.15 + 0.85 = 20$ ft., and

$$\text{the average velocity} = \tfrac{20}{10} = 2 \text{ ft./sec.}$$

(3) Consider a similar problem when the distance is given by $x = 0.2t^2 - 3t + 10.4$. Then $v = 0.4t - 3$, $a = 0.4 = \text{constant}$. When $t = 0$, $x = 10.4$ as before, and $v = -3$; so that the body was moving in the direction opposite to its motion in the previous

cases. As the acceleration is positive, however, we see that this velocity will decrease as time goes on, until it becomes zero, when $v = 0$ or $0.4t - 3 = 0$; or $t = +7.5$ sec. After this, the velocity becomes positive; and 5 seconds after the body started, $t = 12.5$, and

$$v = 0.4 \times 12.5 - 3 = 2 \text{ ft./sec.}$$

When $x = 100$,

$$100 = 0.2t^2 - 3t + 10.4, \quad \text{or} \quad t^2 - 15t - 448 = 0$$

and $\quad t = 29.96;\ v = 0.4 \times 29.96 - 3 = 8.98$ ft./sec.

When v is zero, $x = 0.2 \times 7.5^2 - 3 \times 7.5 + 10.4 = -0.85$, informing us that the body moves back to 0.85 ft. beyond the point O before it stops. Ten seconds later $t = 17.5$ and

$$x = 0.2 \times 17.5^2 - 3 \times 17.5 + 10.4 = 19.15$$

The distance travelled $= 0.85 + 19.15 = 20.0$, and the average velocity is again 2 ft./sec.

(4) Consider yet another problem of the same sort with $x = 0.2t^3 - 3t^2 + 10.4$; $v = 0.6t^2 - 6t$; $a = 1.2t - 6$. The acceleration is no more constant.

When $t = 0$, $x = 10.4$, $v = 0$, $a = -6$. The body is at rest, but just ready to move with a negative acceleration, that is, to gain a velocity towards the point O.

(5) If we have $x = 0.2t^3 - 3t + 10.4$, then $v = 0.6t^2 - 3$, and $a = 1.2t$.

When $t = 0$, $x = 10.4$; $v = -3$; $a = 0$.

The body is moving towards the point O with a velocity of 3 ft./sec., and just at that instant the velocity is uniform.

We see that the conditions of the motion can always be at once ascertained from the time-distance equation and its first and second derived functions. In the last two cases the mean velocity during the first 10 seconds and the velocity 5 seconds after the start will no more be the same, because the velocity is not increasing uniformly, the acceleration being no longer constant.

(6) The angle θ (in radians) turned through by a wheel is given by $\theta = 3 + 2t - 0.1t^3$, where t is the time in seconds from a certain instant; find the angular velocity ω and the angular accelera-

tion α, (*a*) after 1 second; (*b*) after it has performed one revolution. At what time is it at rest, and how many revolutions has it performed up to that instant?

$$\omega = \dot{\theta} = \frac{d\theta}{dt} = 2 - 0.3t^2, \quad \alpha = \ddot{\theta} = \frac{d^2\theta}{dt^2} = -0.6t$$

When $t = 0$, $\theta = 3$; $\omega = 2$ rad./sec.; $\alpha = 0$.
When $t = 1$, $\omega = 2 - 0.3 = 1.7$ rad./sec.; $\alpha = -0.6$ rad./sec^2.
This is a retardation; the wheel is slowing down.
After 1 revolution

$$\theta = 2\pi = 3 + 2t - 0.1t^3$$

By solving this equation numerically we can get the value or values of t for which $\theta = 2\pi$; these are about 2.11 and 3.02 (there is a third negative value).

When $t = 2.11$,

$$\theta = 6.28; \; \omega = 2 - 1.34 = 0.66 \text{ rad./sec.}$$

$$\alpha = -1.27 \text{ rad./sec}^2$$

When $t = 3.02$,

$$\theta = 6.28; \; \omega = 2 - 2.74 = -0.74 \text{ rad./sec.}$$

$$\alpha = -1.81 \text{ rad./sec}^2$$

The velocity is reversed. The wheel is evidently at rest between these two instants; it is at rest when $\omega = 0$, that is when $0 = 2 - 0.3t^2$, or when $t = 2.58$ sec., it has performed

$$\frac{\theta}{2\pi} = \frac{3 + 2 \times 2.58 - 0.1 \times 2.58^3}{6.28} = 1.025 \text{ revolutions}$$

EXERCISES V

(1) If $y = a + bt^2 + ct^4$; find $\frac{dy}{dt}$ and $\frac{d^2y}{dt^2}$.

Ans. $\frac{dy}{dt} = 2bt + 4ct^3$; $\frac{d^2y}{dt^2} = 2b + 12ct^2$

(2) A body falling freely in space describes in t seconds a space s, in feet, expressed by the equation $s = 16t^2$. Draw a curve showing the relation between s and t. Also determine the velocity of the body at the following times from its being let drop: $t = 2$ seconds; $t = 4.6$ seconds; $t = 0.01$ second.[1]

(3) If $x = at - \frac{1}{2}gt^2$; find $\dot{x}$ and $\ddot{x}$.

(4) If a body moves according to the law

$$s = 12 - 4.5t + 6.2t^2$$

find its velocity when $t = 4$ seconds; s being in feet.

(5) Find the acceleration of the body mentioned in the preceding example. Is the acceleration the same for all values of t?

(6) The angle θ (in radians) turned through by a revolving wheel is connected with the time t (in seconds) that has elapsed since starting, by the law

$$\theta = 2.1 - 3.2t + 4.8t^2$$

Find the angular velocity (in radians per second) of that wheel when $1\frac{1}{2}$ seconds have elapsed. Find also its angular acceleration.

1. It is good to be clear on just how derivatives apply to falling bodies, the most familiar instance of accelerated motion. Let t be the time in seconds from the moment a stone is dropped, and s the distance it has fallen. The function relating these two variables is $s = 16t^2$. (It graphs as a neat parabola.) Thus after one second the stone has fallen 16 feet, after two seconds it has dropped $4 \times 16 = 64$ feet, after three seconds, $9 \times 16 = 144$ feet, and so on. The first derivative is $32t$. This gives the instantaneous velocity at which the stone is falling at the end of t seconds. After the first second its velocity is 32 feet per second. After two seconds its velocity is 64 feet per second, and so on.

The second derivative is simply 32. This derivative of a derivative of a function" is the stone's acceleration—the rate at which its velocity is increasing. Physicists call it the "gravitation constant" for bodies falling to the earth's surface.

Later on, in chapters on integration, you will see how integrating the falling stone's first derivative will give the distance traveled by the stone between any two moments as it goes from the start of its fall until the time it is stopped by the ground.—M.G.

(7) A slider moves so that, during the first part of its motion, its distance s in inches from its starting point is given by the expression

$$s = 6.8t^3 - 10.8t; \; t \text{ being in seconds.}$$

Find the expression for the velocity and the acceleration at any time; and hence find the velocity and the acceleration after 3 seconds.

(8) The motion of a rising balloon is such that its height h, in miles, is given at any instant by the expression

$$h = 0.5 + \tfrac{1}{10}\sqrt[3]{t - 125}; \; t \text{ being in seconds.}$$

Find an expression for the velocity and the acceleration at any time. Draw curves to show the variation of height, velocity and acceleration during the first ten minutes of the ascent.

(9) A stone is thrown downwards into water and its depth p in meters at any instant t seconds after reaching the surface of the water is given by the expression

$$p = \frac{4}{4 + t^2} + 0.8t - 1$$

Find an expression for the velocity and the acceleration at any time. Find the velocity and acceleration after 10 seconds.

(10) A body moves in such a way that the space described in the time t from starting is given by $s = t^n$, where n is a constant. Find the value of n when the velocity is doubled from the 5th to the 10th second; find it also when the velocity is numerically equal to the acceleration at the end of the 10th second.

Chapter IX

INTRODUCING A USEFUL DODGE

Sometimes one is stumped by finding that the expression to be differentiated is too complicated to tackle directly.

Thus, the equation

$$y = (x^2 + a^2)^{\frac{3}{2}}$$

is awkward to a beginner.

Now the dodge to turn the difficulty is this: Write some symbol, such as u, for the expression $x^2 + a^2$; then the equation becomes

$$y = u^{\frac{3}{2}}$$

which you can easily manage; for

$$\frac{dy}{du} = \frac{3}{2}u^{\frac{1}{2}}$$

Then tackle the expression

$$u = x^2 + a^2$$

and differentiate it with respect to x

$$\frac{du}{dx} = 2x$$

Then all that remains is plain sailing;

for $$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

that is,
$$\frac{dy}{dx} = \frac{3}{2}u^{\frac{1}{2}} \times 2x$$
$$= \tfrac{3}{2}(x^2 + a^2)^{\frac{1}{2}} \times 2x$$
$$= 3x(x^2 + a^2)^{\frac{1}{2}}$$

and so the trick is done.[1]

By and by, when you have learned how to deal with sines, and cosines, and exponentials, you will find this dodge of increasing usefulness.

Examples.

Let us practice this dodge on a few examples.

(1) Differentiate $y = \sqrt{a + x}$.

Let $u = a + x$.

$$\frac{du}{dx} = 1;\ y = u^{\frac{1}{2}};\ \frac{dy}{du} = \tfrac{1}{2}u^{-\frac{1}{2}} = \tfrac{1}{2}(a + x)^{-\frac{1}{2}}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx} = \frac{1}{2\sqrt{a + x}}$$

1. Thompson's "useful dodge" is known today as the "chain rule." It is one of the most useful rules in calculus. The function he gives to illustrate it is called a "composite function" because it involves a "function of a function." The expression inside the parentheses $(x^2 + a^2)$ is called the "inside function." The "outside function" is the exponent 3/2.

One could try to differentiate by cubing $(x^2 + a^2)$ and taking its square root, or by expanding $(x^2 + a^2)^{\frac{3}{2}}$ by the binomial theorem, but neither of these methods works very well. As Thompson makes clear, the much simpler way is to differentiate the outside function with respect to the inside one, then multiply the result by the derivative of the inside function with respect to x. It is called the chain rule because it can be applied to composite functions with more than one inside function. You simply compute a chain of derivatives, then multiply them together. Modern calculus texts give proof of why the chain rule works.

For a trivial example of how it works, consider three children A, B, and C. A grows twice as fast as B, and B grows three times as fast as C. How much faster is A growing than C? Clearly the answer is $2 \times 3 = 6$ times as fast.—M.G.

(2) Differentiate $y = \dfrac{1}{\sqrt{a+x^2}}$

Let $u = a + x^2$.

$$\frac{du}{dx} = 2x;\ y = u^{-\frac{1}{2}};\ \frac{dy}{du} = -\tfrac{1}{2}u^{-\frac{3}{2}}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx} = -\frac{x}{\sqrt{(a+x^2)^3}}$$

(3) Differentiate $y = \left(m - nx^{\frac{2}{3}} + \dfrac{p}{x^{\frac{4}{3}}}\right)^a$

Let $u = m - nx^{\frac{2}{3}} + px^{-\frac{4}{3}}$.

$$\frac{du}{dx} = -\tfrac{2}{3}nx^{-\frac{1}{3}} - \tfrac{4}{3}px^{-\frac{7}{3}}$$

$$y = u^a;\ \frac{dy}{du} = au^{a-1}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx} = -a\left(m - nx^{\frac{2}{3}} + \frac{p}{x^{\frac{4}{3}}}\right)^{a-1} \left(\tfrac{2}{3}nx^{-\frac{1}{3}} + \tfrac{4}{3}px^{-\frac{7}{3}}\right)$$

(4) Differentiate $y = \dfrac{1}{\sqrt{x^3 - a^2}}$

Let $u = x^3 - a^2$.

$$\frac{du}{dx} = 3x^2;\ y = u^{-\frac{1}{2}};\ \frac{dy}{du} = -\tfrac{1}{2}(x^3 - a^2)^{-\frac{3}{2}}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx} = -\frac{3x^2}{2\sqrt{(x^3 - a^2)^3}}$$

(5) Differentiate $y = \sqrt{\dfrac{1-x}{1+x}}$

Write this as $\qquad y = \dfrac{(1-x)^{\frac{1}{2}}}{(1+x)^{\frac{1}{2}}}$

$$\frac{dy}{dx}=\frac{(1+x)^{\frac{1}{2}}\dfrac{d(1-x)^{\frac{1}{2}}}{dx}-(1-x)^{\frac{1}{2}}\dfrac{d(1+x)^{\frac{1}{2}}}{dx}}{1+x}$$

(We may also write $y=(1-x)^{\frac{1}{2}}(1+x)^{-\frac{1}{2}}$ and differentiate as a product.)

Proceeding as in Example (1) above, we get

$$\frac{d(1-x)^{\frac{1}{2}}}{dx}=-\frac{1}{2\sqrt{1-x}};\quad\text{and}\quad\frac{d(1+x)^{\frac{1}{2}}}{dx}=\frac{1}{2\sqrt{1+x}}$$

Hence $$\frac{dy}{dx}=-\frac{(1+x)^{\frac{1}{2}}}{2(1+x)\sqrt{1-x}}-\frac{(1-x)^{\frac{1}{2}}}{2(1+x)\sqrt{1+x}}$$

$$=-\frac{1}{2\sqrt{1+x}\sqrt{1-x}}-\frac{\sqrt{1-x}}{2\sqrt{(1+x)^3}}$$

or $$\frac{dy}{dx}=-\frac{1}{(1+x)\sqrt{1-x^2}}$$

(6) Differentiate $y=\sqrt{\dfrac{x^3}{1+x^2}}$

We may write this

$$y=x^{\frac{3}{2}}(1+x^2)^{-\frac{1}{2}}$$

$$\frac{dy}{dx}=\tfrac{3}{2}x^{\frac{1}{2}}(1+x^2)^{-\frac{1}{2}}+x^{\frac{3}{2}}\times\frac{d\left[(1+x^2)^{-\frac{1}{2}}\right]}{dx}$$

Differentiating $(1+x^2)^{-\frac{1}{2}}$, as shown in Example (2) above, we get

$$\frac{d\left[(1+x^2)^{-\frac{1}{2}}\right]}{dx}=-\frac{x}{\sqrt{(1+x^2)^3}}$$

so that $$\frac{dy}{dx}=\frac{3\sqrt{x}}{2\sqrt{1+x^2}}-\frac{\sqrt{x^5}}{\sqrt{(1+x^2)^3}}=\frac{\sqrt{x}(3+x^2)}{2\sqrt{(1+x^2)^3}}$$

(7) Differentiate $y=\left(x+\sqrt{x^2+x+a}\right)^3$

Let $u=x+\sqrt{x^2+x+a}$

$$\frac{du}{dx} = 1 + \frac{d\left[(x^2+x+a)^{\frac{1}{2}}\right]}{dx}$$

$$y = u^3; \quad \text{and} \quad \frac{dy}{du} = 3u^2 = 3\left(x + \sqrt{x^2+x+a}\right)^2$$

Now let $v = (x^2+x+a)^{\frac{1}{2}}$ and $w = (x^2+x+a)$

$$\frac{dw}{dx} = 2x+1; \; v = w^{\frac{1}{2}}; \; \frac{dv}{dw} = \tfrac{1}{2}w^{-\frac{1}{2}}$$

$$\frac{dv}{dx} = \frac{dv}{dw} \times \frac{dw}{dx} = \tfrac{1}{2}(x^2+x+a)^{-\frac{1}{2}}(2x+1)$$

Hence $$\frac{du}{dx} = 1 + \frac{2x+1}{2\sqrt{x^2+x+a}}$$

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

$$= 3\left(x+\sqrt{x^2+x+a}\right)^2\left(1 + \frac{2x+1}{2\sqrt{x^2+x+a}}\right)$$

(8) Differentiate $y = \sqrt{\dfrac{a^2+x^2}{a^2-x^2}}\;\sqrt[3]{\dfrac{a^2-x^2}{a^2+x^2}}$

We get $$y = \frac{(a^2+x^2)^{\frac{1}{2}}(a^2-x^2)^{\frac{1}{3}}}{(a^2-x^2)^{\frac{1}{2}}(a^2+x^2)^{\frac{1}{3}}} = (a^2+x^2)^{\frac{1}{6}}\,(a^2-x^2)^{-\frac{1}{6}}$$

$$\frac{dy}{dx} = (a^2+x^2)^{\frac{1}{6}}\,\frac{d\left[(a^2-x^2)^{-\frac{1}{6}}\right]}{dx} + \frac{d\left[(a^2+x^2)^{\frac{1}{6}}\right]}{(a^2-x^2)^{\frac{1}{6}}dx}$$

Let $u = (a^2-x^2)^{-\frac{1}{6}}$ and $v = (a^2-x^2)$

$$u = v^{-\frac{1}{6}}; \; \frac{du}{dv} = -\frac{1}{6}v^{-\frac{7}{6}}; \; \frac{dv}{dx} = -2x$$

$$\frac{du}{dx} = \frac{du}{dv} \times \frac{dv}{dx} = \frac{1}{3}x(a^2-x^2)^{-\frac{7}{6}}$$

Let $w = (a^2 + x^2)^{\frac{1}{6}}$ and $z = (a^2 + x^2)$

$$w = z^{\frac{1}{6}};\ \frac{dw}{dz} = \frac{1}{6} z^{-\frac{5}{6}};\ \frac{dz}{dx} = 2x$$

$$\frac{dw}{dx} = \frac{dw}{dz} \times \frac{dz}{dx} = \frac{1}{3} x(a^2 + x^2)^{-\frac{5}{6}}$$

Hence $$\frac{dy}{dx} = (a^2 + x^2)^{\frac{1}{6}} \frac{x}{3(a^2 - x^2)^{\frac{7}{6}}} + \frac{x}{3(a^2 - x^2)^{\frac{1}{6}}(a^2 + x^2)^{\frac{5}{6}}}$$

or $$\frac{dy}{dx} = \frac{x}{3}\left[\sqrt[6]{\frac{a^2 + x^2}{(a^2 - x^2)^7}} + \frac{1}{\sqrt[6]{(a^2 - x^2)(a^2 + x^2)^5}}\right]$$

(9) Differentiate y^n with respect to y^5.

$$\frac{d(y^n)}{d(y^5)} = \frac{ny^{n-1}}{5y^{5-1}} = \frac{n}{5} y^{n-5}$$

(10) Find the first and second derivatives of $y = \frac{x}{b}\sqrt{(a - x)x}$.

$$\frac{dy}{dx} = \frac{x}{b} \frac{d\left\{[(a - x)x]^{\frac{1}{2}}\right\}}{dx} + \frac{\sqrt{(a - x)x}}{b}$$

Let $u = [(a - x)x]^{\frac{1}{2}}$ and let $w = (a - x)x$; then $u = w^{\frac{1}{2}}$

$$\frac{du}{dw} = \frac{1}{2} w^{-\frac{1}{2}} = \frac{1}{2w^{\frac{1}{2}}} = \frac{1}{2\sqrt{(a - x)x}}$$

$$\frac{dw}{dx} = a - 2x$$

$$\frac{du}{dw} \times \frac{dw}{dx} = \frac{du}{dx} = \frac{a - 2x}{2\sqrt{(a - x)x}}$$

Hence $$\frac{dy}{dx} = \frac{x(a - 2x)}{2b\sqrt{(a - x)x}} + \frac{\sqrt{(a - x)x}}{b} = \frac{x(3a - 4x)}{2b\sqrt{(a - x)x}}$$

Now $$\frac{d^2y}{dx^2} = \frac{2b\sqrt{(a-x)x}(3a-8x) - \dfrac{(3ax-4x^2)b(a-2x)}{\sqrt{(a-x)x}}}{4b^2(a-x)x}$$

$$= \frac{3a^2 - 12ax + 8x^2}{4b(a-x)\sqrt{(a-x)x}}$$

(We shall need these two last derivatives later on. See Ex. X, No. 11.)

EXERCISES VI

Differentiate the following:

(1) $y = \sqrt{x^2+1}$ (2) $y = \sqrt{x^2+a^2}$

(3) $y = \dfrac{1}{\sqrt{a+x}}$ (4) $y = \dfrac{a}{\sqrt{a-x^2}}$ (5) $y = \dfrac{\sqrt{x^2-a^2}}{x^2}$

(6) $y = \dfrac{\sqrt[3]{x^4+a}}{\sqrt[2]{x^3+a}}$ (7) $y = \dfrac{a^2+x^2}{(a+x)^2}$

(8) Differentiate y^5 with respect to y^2

(9) Differentiate $y = \dfrac{\sqrt{1-\theta^2}}{1-\theta}$

The process can be extended to three or more derivatives, so that $\dfrac{dy}{dx} = \dfrac{dy}{dz} \times \dfrac{dz}{dv} \times \dfrac{dv}{dx}$

Examples.

(1) If $z = 3x^4$; $v = \dfrac{7}{z^2}$; $y = \sqrt{1+v}$, find $\dfrac{dy}{dx}$

We have $\dfrac{dy}{dv} = \dfrac{1}{2\sqrt{1+v}}$; $\dfrac{dv}{dz} = -\dfrac{14}{z^3}$; $\dfrac{dz}{dx} = 12x^3$

$$\frac{dy}{dx} = -\frac{168x^3}{\left(2\sqrt{1+v}\right)z^3} = -\frac{28}{3x^5\sqrt{9x^8+7}}$$

(2) If $t = \dfrac{1}{5\sqrt{\theta}}$; $x = t^3 + \dfrac{t}{2}$; $v = \dfrac{7x^2}{\sqrt[3]{x-1}}$, find $\dfrac{dv}{d\theta}$

$$\frac{dv}{dx} = \frac{7x(5x-6)}{3\sqrt[3]{(x-1)^4}};\ \frac{dx}{dt} = 3t^2 + \tfrac{1}{2};\ \frac{dt}{d\theta} = -\frac{1}{10\sqrt{\theta^3}}$$

Hence
$$\frac{dv}{d\theta} = -\frac{7x(5x-6)\left(3t^2+\frac{1}{2}\right)}{30\sqrt[3]{(x-1)^4}\sqrt{\theta^3}}$$

an expression in which x must be replaced by its value, and t by its value in terms of θ.

(3) If $\theta = \dfrac{3a^2x}{\sqrt{x^3}}$; $\omega = \dfrac{\sqrt{1-\theta^2}}{1+\theta}$ and $\phi = \sqrt{3} - \dfrac{1}{\omega\sqrt{2}}$, find $\dfrac{d\phi}{dx}$.

We get $\theta = 3a^2x^{-\frac{1}{2}}$; $\omega = \sqrt{\dfrac{1-\theta}{1+\theta}}$; and $\phi = \sqrt{3} - \dfrac{1}{\sqrt{2}}\omega^{-1}$

$$\frac{d\theta}{dx} = -\frac{3a^2}{2\sqrt{x^3}};\ \frac{d\omega}{d\theta} = -\frac{1}{(1+\theta)\sqrt{1-\theta^2}}$$

(see example 5, Capter IX); and

$$\frac{d\phi}{d\omega} = \frac{1}{\sqrt{2}\cdot\omega^2}$$

So that
$$\frac{d\phi}{dx} = \frac{1}{\sqrt{2}\times\omega^2} \times \frac{1}{(1+\theta)\sqrt{1-\theta^2}} \times \frac{3a^2}{2\sqrt{x^3}}$$

Replace now first ω, then θ by its value.

EXERCISES VII

You can now successfully try the following.

(1) If $u = \frac{1}{2}x^3$; $v = 3(u+u^2)$; and $w = \dfrac{1}{v^2}$, find $\dfrac{dw}{dx}$

(2) If $y = 3x^2 + \sqrt{2}$; $z = \sqrt{1 + y}$; and $v = \dfrac{1}{\sqrt{3} + 4z}$, find $\dfrac{dv}{dx}$

(3) If $y = \dfrac{x^3}{\sqrt{3}}$; $z = (1 + y)^2$; and $u = \dfrac{1}{\sqrt{1 + z}}$, find $\dfrac{du}{dx}$

The following exercises are placed here for reasons of space and because their solution depends upon the dodge explained in the foregoing chapter, but they should not be attempted until Chapters XIV and XV have been read.

(4) If $y = 2a^3 \log_e u - u\left(5a^2 - 2au + \frac{1}{3}u^2\right)$, and $u = a + x$,

show that
$$\frac{dy}{dx} = \frac{x^2(a - x)}{a + x}$$

(5) For the curve $x = a(\theta - \sin\theta)$, $y = a(1 - \cos\theta)$, find $\dfrac{dx}{d\theta}$ and $\dfrac{dy}{d\theta}$; hence deduce the value of $\dfrac{dy}{dx}$

(6) Find $\dfrac{dx}{d\theta}$ and $\dfrac{dy}{d\theta}$ for the curve $x = a\cos^3\theta$, $y = a\sin^3\theta$; hence obtain $\dfrac{dy}{dx}$

(7) Given that $y = \log_e \sin(x^2 - a^2)$, find $\dfrac{dy}{dx}$ in its simplest form.

(8) If $u = x + y$ and $4x = 2u - \log_e(2u - 1)$, show that
$$\frac{dy}{dx} = \frac{x + y}{x + y - 1}$$

Chapter X

GEOMETRICAL MEANING OF DIFFERENTIATION

It is useful to consider what geometrical meaning can be given to the derivative.

In the first place, any function of x, such, for example, as x^2, or $\sqrt{x}$, or $ax + b$, can be plotted as a curve; and nowadays every schoolboy is familiar with the process of curve plotting.

Let PQR, in Fig. 7, be a portion of a curve plotted with respect to the axes of coordinates OX and OY. Consider any point Q on this curve, where the abscissa of the point is x and its ordinate is y. Now observe how y changes when x is varied. If x is made to increase by a small increment dx, to the right, it will be observed that y also (in *this* particular curve) increases by a small increment dy (because this particular curve happens to be an *ascending* curve). Then the ratio of dy to dx is a measure of the degree to which the curve is sloping up between the two points Q and T. As a matter of fact, it can be seen on the figure that the curve between Q and T has many different slopes, so that we cannot very well speak of the slope of the curve between Q and T. If, however, Q and T are so near each other that the small portion QT of the curve is practically straight, then it is true to say that the ratio $\dfrac{dy}{dx}$ is the slope of the curve along QT. The straight line QT produced on either side touches the curve along the portion QT only, and if this portion is infinitely small, the straight

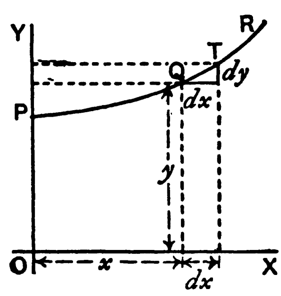

FIG. 7.

line will touch the curve at practically one point only, and be therefore a *tangent* to the curve.

This tangent to the curve has evidently the same slope as QT, so that $\frac{dy}{dx}$ is the slope of the tangent to the curve at the point Q for which the value of $\frac{dy}{dx}$ is found.

We have seen that the short expression "the slope of a curve" has no precise meaning, because a curve has so many slopes—in fact, every small portion of a curve has a different slope. "The slope of a curve *at a point*" is, however, a perfectly defined thing; it is the slope of a very small portion of the curve situated just at that point; and we have seen that this is the same as "the slope of the tangent to the curve at that point".

Observe that dx is a short step to the right, and dy the corresponding short step upwards. These steps must be considered as short as possible—in fact infinitely short,—though in diagrams we have to represent them by bits that are not infinitesimally small, otherwise they could not be seen.

We shall hereafter make considerable use of this circumstance that $\frac{dy}{dx}$ *represents the slope of the tangent to the curve at any point.*

If a curve is sloping up at 45° at a particular point, as in Fig. 8, dy and dx will be equal, and the value of $\frac{dy}{dx} = 1$.

If the curve slopes up steeper than 45° (Fig. 9), $\frac{dy}{dx}$ will be greater than 1.

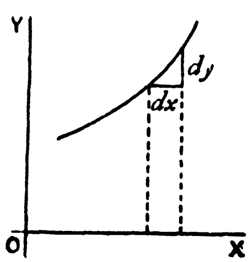

FIG. 8.

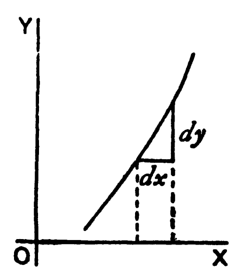

FIG. 9.

If the curve slopes up very gently, as in Fig. 10, $\frac{dy}{dx}$ will be a fraction smaller than 1.

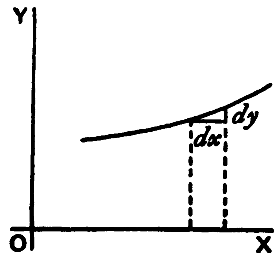

FIG. 10.

For a horizontal line, or a horizontal place in a curve, $dy = 0$, and therefore $\frac{dy}{dx} = 0$.

If a curve slopes *downward*, as in Fig. 11, dy will be a step down, and must therefore be reckoned of negative value; hence $\frac{dy}{dx}$ will have negative sign also.

If the "curve" happens to be a straight line, like that in Fig. 12, the value of $\frac{dy}{dx}$ will be the same at all points along it. In other words its *slope* is constant.

If a curve is one that turns more upwards as it goes along to the right, the values of $\frac{dy}{dx}$ will become greater and greater with the increasing steepness, as in Fig. 13.

If a curve is one that gets flatter and flatter as it goes along, the values of $\frac{dy}{dx}$ will become smaller and smaller as the flatter part is reached, as in Fig. 14.

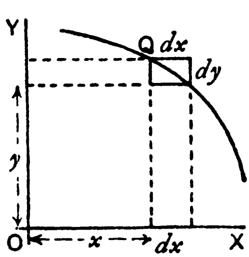

FIG. 11.

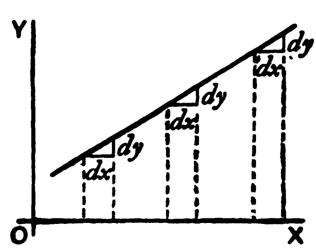

FIG. 12.

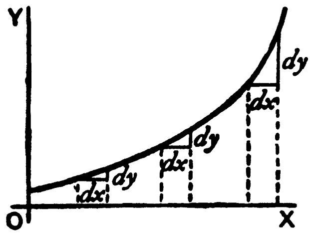

FIG. 13.

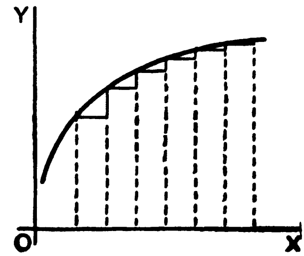

FIG. 14.

If a curve first descends, and then goes up again, as in Fig. 15, presenting a concavity upwards, then clearly $\dfrac{dy}{dx}$ will first be negative, with diminishing values as the curve flattens, then will be zero at the point where the bottom of the trough of the curve is reached; and from this point onward $\dfrac{dy}{dx}$ will have positive values that go on increasing. In such a case y is said to pass through a *local minimum*. This value of y is not necessarily the smallest value of y, it is that value of y corresponding to the bottom of the trough; for instance, in Fig. 28, the value of y corresponding to the bottom of the trough is 1, while y takes elsewhere values which are smaller than this. The characteristic of a local minimum is that y must increase *on either side* of it.

N.B.—For the particular value of x that makes y *a minimum*, the value of $\dfrac{dy}{dx} = 0$.

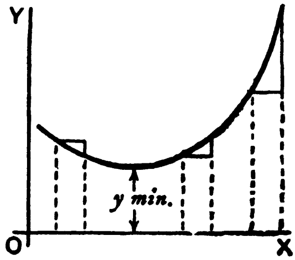

FIG. 15.

If a curve first ascends and then descends, the values of $\dfrac{dy}{dx}$ will be positive at first; then zero, as the summit is reached; then negative, as the curve slopes downwards, as in Fig. 16.

In this case y is said to pass through a *local maximum*, but this value of y is

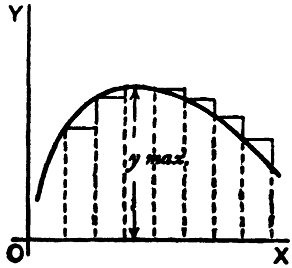

FIG. 16.

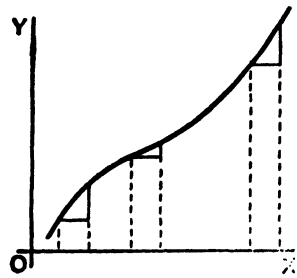

FIG. 17.

not necessarily the greatest value of y. In Fig. 28, the local maximum of y is $2\frac{1}{3}$, but this is by no means the greatest value y can have at some other point of the curve.

N.B.—For the particular value of x that makes y *a maximum*, the value of $\dfrac{dy}{dx} = 0$.

If a curve has the particular form of Fig. 17, the values of $\dfrac{dy}{dx}$ will always be positive; but there will be one particular place where the slope is least steep, where the value of $\dfrac{dy}{dx}$ will be a minimum; that is, less than it is at any other part of the curve.

If a curve has the form of Fig. 18, the value of $\dfrac{dy}{dx}$ will be negative in the upper part, and positive in the lower part; while at the nose of the curve where it becomes actually perpendicular, the value of $\dfrac{dy}{dx}$ will be infinitely great.

Now that we understand that $\dfrac{dy}{dx}$ measures the steepness of a curve at any point, let us turn to some of the equations which we have already learned how to differentiate.

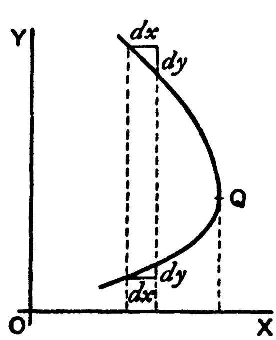

FIG. 18.

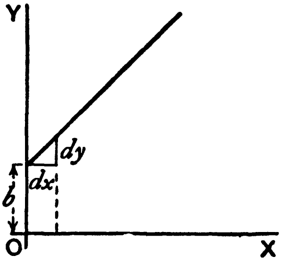

FIG. 19.

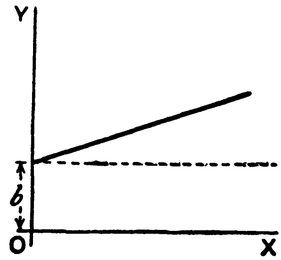

FIG. 20.

(1) As the simplest case take this:

$$y = x + b$$

It is plotted out in Fig. 19, using equal scales for x and y. If we put $x = 0$, then the corresponding ordinate will be $y = b$; that is to say, the "curve" crosses the y-axis at the height b. From here it ascends at 45°; for whatever values we give to x to the right, we have an equal y to ascend. The line has a gradient of 1 in 1.

Now differentiate $y = x + b$, by the rules we have already learned and we get $\dfrac{dy}{dx} = 1$.

The slope of the line is such that for every little step dx to the right, we go an equal little step dy upward. And this slope is constant—always the same slope.

(2) Take another case:

$$y = ax + b.$$

We know that this curve, like the preceding one, will start from a height b on the y-axis. But before we draw the curve, let us find its slope by differentiating; which gives us $\dfrac{dy}{dx} = a$. The slope will be constant, at an angle, the tangent of which is here called a. Let us assign to a some numerical value—say $\frac{1}{3}$. Then we must give it such a slope that it ascends 1 in 3; or dx will be 3 times as great as dy; as magnified in Fig. 21. So, draw the line in Fig. 20 at this slope.

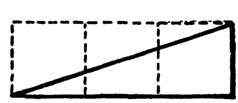

FIG. 21.

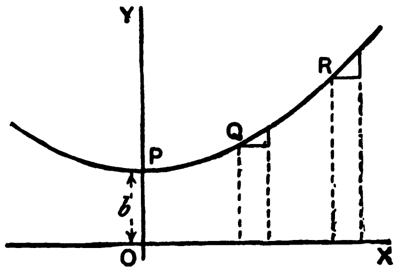

FIG. 22.

(3) Now for a slightly harder case. Let

$$y = ax^2 + b$$

Again the curve will start on the y-axis at a height b above the origin.

Now differentiate. [If you have forgotten, turn back to Chapter V, or, rather, *don't* turn back, but think out the differentiation.]

$$\frac{dy}{dx} = 2ax$$

This shows that the steepness will not be constant: it increases as x increases. At the starting point P, where $x = 0$, the curve (Fig. 22) has no steepness—that is, it is level. On the left of the origin, where x has negative values, $\frac{dy}{dx}$ will also have negative values, or will descend from left to right, as in the Figure.

Let us illustrate this by working out a particular instance. Taking the equation

$$y = \tfrac{1}{4}x^2 + 3$$

and differentiating it, we get

$$\frac{dy}{dx} = \tfrac{1}{2}x$$

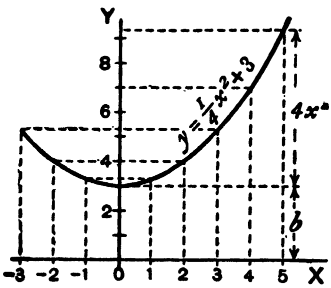

FIG. 23.

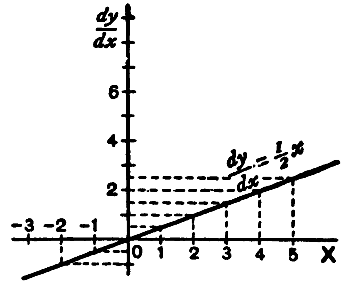

FIG. 24.

Now assign a few successive values, say from 0 to 5, to x; and calculate the corresponding values of y by the first equation; and of $\frac{dy}{dx}$ from the second equation. Tabulating results, we have:

x	0	1	2	3	4	5
y	3	$3\frac{1}{4}$	4	$5\frac{1}{4}$	7	$9\frac{1}{4}$
$\frac{dy}{dx}$	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$

Then plot them out in two curves, Figs. 23 and 24; in Fig. 23 plotting the values of y against those of x, and in Fig. 24 those of $\frac{dy}{dx}$ against those of x. For any assigned value of x, the *height* of the ordinate in the second curve is proportional to the *slope* of the first curve.

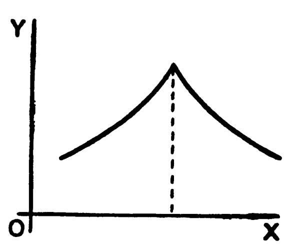

FIG. 25.

If a curve comes to a sudden *cusp,* as in Fig. 25, the slope at that point suddenly changes from a slope upward to a slope downward. In that case $\frac{dy}{dx}$

will clearly undergo an abrupt change from a positive to a negative value.[1]

The following examples show further applications of the principles just explained.

(4) Find the slope of the tangent to the curve $y = \frac{1}{2x} + 3$ at the point where $x = -1$. Find the angle which this tangent makes with the curve $y = 2x^2 + 2$.

The slope of the tangent is the slope of the curve at the point where they touch one another; that is, it is the $\frac{dy}{dx}$ of the curve for that point. Here $\frac{dy}{dx} = -\frac{1}{2x^2}$ and for $x = -1$, $\frac{dy}{dx} = -\frac{1}{2}$, which is the slope of the tangent and of the curve at that point. The tangent, being a straight line, has for equation $y = ax + b$, and its slope is $\frac{dy}{dx} = a$, hence $a = -\frac{1}{2}$. Also if $x = -1$, $y = \frac{1}{2(-1)} + 3 = 2\tfrac{1}{2}$; and as the tangent passes by this point, the coordinates of the point must satisfy the equation of the tangent, namely

$$y = -\frac{1}{2}x + b$$

so that $2\tfrac{1}{2} = -\frac{1}{2} \times (-1) + b$ and $b = 2$; the equation of the tangent is therefore $y = -\frac{1}{2}x + 2$

1. A cusp is the sharp point on a curve where the curve abruptly changes its direction by 180°. If it changes by some other angle the point is called a "corner." A cusp or corner can, of course, point north as well as south, east, or west, or in any other direction. The tangent at the cusp shown in Figure 25 is vertical. The curve approaches the tangent from one side, leaves it on the other side. At the cusp point there is no derivative. A typical example is the curve for $y = (x - 4)^{\frac{2}{3}}$. Its cusp points south at $x = 4$.

Corners can occur on graphs of continuous functions that are made up of straight lines. For example, the absolute value function, defined to be x if $x \geq 0$ and $-x$ if $x < 0$, has a corner at $x = 0$.—M.G.

Now, when two curves meet, the intersection being a point common to both curves, its coordinates must satisfy the equation of each one of the two curves; that is, it must be a solution of the system of simultaneous equations formed by coupling together the equations of the curves. Here the curves meet one another at points given by the solution of

$$\begin{cases} y = 2x^2 + 2 \\ y = -\frac{1}{2}x + 2 \end{cases} \quad \text{or} \quad 2x^2 + 2 = -\tfrac{1}{2}x + 2$$

that is,
$$x\left(2x + \tfrac{1}{2}\right) = 0$$

This equation has for its solutions $x = 0$ and $x = -\frac{1}{4}$. The slope of the curve $y = 2x^2 + 2$ at any point is

$$\frac{dy}{dx} = 4x$$

For the point where $x = 0$, this slope is zero; the curve is horizontal. For the point where

$$x = -\frac{1}{4}, \quad \frac{dy}{dx} = -1$$

hence the curve at that point slopes downwards to the right at such an angle θ with the horizontal that $\tan\theta = 1$; that is, at $45°$ to the horizontal.

The slope of the straight line is $-\frac{1}{2}$; that is, it slopes downwards to the right and makes with the horizontal an angle ϕ such that $\tan\phi = \frac{1}{2}$; that is, an angle of $26°\ 34'$. It follows that at the first point the curve cuts the straight line at an angle of $26°\ 34'$, while at the second it cuts it at an angle of $45° - 26°\ 34' = 18°\ 26'$.

(5) A straight line is to be drawn, through a point whose coordinates are $x = 2$, $y = -1$, as tangent to the curve

$$y = x^2 - 5x + 6$$

Find the coordinates of the point of contact.

The slope of the tangent must be the same as the $\dfrac{dy}{dx}$ of the curve; that is, $2x - 5$.

The equation of the straight line is $y = ax + b$, and as it is satisfied for the values $x = 2$, $y = -1$, then $-1 = a \times 2 + b$; also, its

$$\frac{dy}{dx} = a = 2x - 5$$

The x and the y of the point of contact must also satisfy both the equation of the tangent and the equation of the curve.

We have then

$$\begin{cases} y = x^2 - 5x + 6, & \text{.............................(i)} \\ y = ax + b, & \text{....................................(ii)} \\ -1 = 2a + b, & \text{...................................(iii)} \\ a = 2x - 5, & \text{...................................(iv)} \end{cases}$$

four equations in a, b, x, y.

Equations (i) and (ii) give $x^2 - 5x + 6 = ax + b$.

Replacing a and b by their value in this, we get

$$x^2 - 5x + 6 = (2x - 5)x - 1 - 2(2x - 5)$$

which simplifies to $x^2 - 4x + 3 = 0$, the solutions of which are: $x = 3$ and $x = 1$. Replacing in (i), we get $y = 0$ and $y = 2$ respectively; the two points of contact are then $x = 1$, $y = 2$; and $x = 3$, $y = 0$.

Note.—In all exercises dealing with curves, students will find it extremely instructive to verify the deductions obtained by actually plotting the curves.

EXERCISES VIII

(1) Plot the curve $y = \frac{3}{4}x^2 - 5$, using a scale of millimeters. Measure at points corresponding to different values of x, the angle of its slope.

Find, by differentiating the equation, the expression for slope; and see from your calculator whether this agrees with the measured angle.

(2) Find what will be the slope of the curve

$$y = 0.12x^3 - 2$$

at the point where $x = 2$.

(3) If $y = (x - a)(x - b)$, show that at the particular point of the curve where $\dfrac{dy}{dx} = 0$, x will have the value $\frac{1}{2}(a + b)$.

(4) Find the $\dfrac{dy}{dx}$ of the equation $y = x^3 + 3x$; and calculate the numerical values of $\dfrac{dy}{dx}$ for the points corresponding to $x = 0$, $x = \frac{1}{2}$, $x = 1$, $x = 2$.

(5) In the curve to which the equation is $x^2 + y^2 = 4$, find the value of x at those points where the slope $= 1$.

(6) Find the slope, at any point, of the curve whose equation is $\dfrac{x^2}{3^2} + \dfrac{y^2}{2^2} = 1$; and give the numerical value of the slope at the place where $x = 0$, and at that where $x = 1$.

(7) The equation of a tangent to the curve $y = 5 - 2x + 0.5x^3$, being of the form $y = mx + n$, where m and n are constants, find the value of m and n if the point where the tangent touches the curve has $x = 2$ for abscissa.

(8) At what angle do the two curves

$$y = 3.5x^2 + 2 \quad \text{and} \quad y = x^2 - 5x + 9.5$$

cut one another?

(9) Tangents to the curve $y = \pm\sqrt{25 - x^2}$ are drawn at points for which $x = 3$ and $x = 4$, the value of y being positive. Find the coordinates of the point of intersection of the tangents and their mutual inclination.

(10) A straight line $y = 2x - b$ touches a curve $y = 3x^2 + 2$ at one point. What are the coordinates of the point of contact, and what is the value of b?

POSTSCRIPT

If the graph of a continuous function crosses the x axis at two points a and b, and is differentiable for the closed interval between a and b, there must be at least one point on the curve between a and b where the tangent is horizontal and the derivative is zero. This is called Rolle's (pronounced Roll's) theorem after French mathematician Michel Rolle (1652-1719).

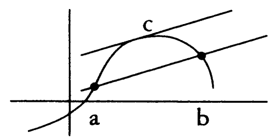

FIG. 25a. Lagrange's mean value theorem.

Rolle's theorem is a special case of what is called Lagrange's mean value theorem, after French mathematician Joseph Louis Lagrange (1736-1813). It states that a straight line between points a and b on the curve for a continuous differentiable function is parallel to the tangent of at least one point c on the closed interval between a and b. (See Figure 25a)

Applied to the velocity of a car that goes at an average speed of, say, 45 miles per hour from A to B, no matter how many times it alters its speed along the way (including even stops), there will be at least one moment when the car's instantaneous speed will be exactly 45 miles per hour.

Both theorems are intuitively obvious, yet they underlie many important, more complicated calculus theorems. For example, the mean value theorem is the basis for calculating antiderivatives.

The Lagrange mean value theorem further generalizes to what is called Cauchy's mean value theorem after Augustin Louis Cauchy (1789-1857), a French mathematician. It concerns two continuous functions that are differentiable in a closed interval. You will learn about it in more advanced calculus textbooks.—M.G.

Chapter XI

MAXIMA AND MINIMA

A quantity which varies continuously is said to pass by (or through) a local maximum or minimum value when, in the course of its variation, the immediately preceding and following values are *both* smaller or greater, respectively, than the value referred to. An infinitely great value is therefore not a maximum value.

One of the principal uses of the process of differentiating is to find out under what conditions the value of the thing differentiated becomes a maximum or a minimum. This is often exceedingly important in engineering and economic questions, where it is most desirable to know what conditions will make the cost of working a minimum, or will make the efficiency a maximum.

Now, to begin with a concrete case, let us take the equation

$$y = x^2 - 4x + 7$$

By assigning a number of successive values to x, and finding the corresponding values of y, we can readily see that the equation represents a curve with a minimum.

x	0	1	2	3	4	5
y	7	4	3	4	7	12

These values are plotted in Fig. 26, which shows that y has apparently a minimum value of 3, when x is made equal to 2. But are you sure that the minimum occurs at 2, and not at $2\frac{1}{4}$ or at $1\frac{3}{4}$?

Of course it would be possible with any algebraic expression to work out a lot of values, and in this way arrive gradually at the particular value that may be a maximum or a minimum.

Here is another example: Let

$$y = 3x - x^2$$

Calculate a few values thus:

x	-1	0	1	2	3	4	5
y	-4	0	2	2	0	-4	-10

Plot these values as in Fig. 27.

It will be evident that there will be a maximum somewhere between $x = 1$ and $x = 2$; and the thing *looks* as if the maximum value of y ought to be about $2\frac{1}{4}$. Try some intermediate values. If $x = 1\frac{1}{4}$, $y = 2.1875$; if $x = 1\frac{1}{2}$, $y = 2.25$; if $x = 1.6$, $y = 2.24$. How can we be sure that 2.25 is the real maximum, or that it occurs exactly when $x = 1\frac{1}{2}$?

Now it may sound like juggling to be assured that there is a way by which one can arrive straight at a maximum (or minimum) value without making a lot of preliminary trials or guesses. And that way depends on differentiating. Look back to Chapter X for the remarks about Figs. 15 and 16, and you will see that whenever a curve gets either to its maximum or to its

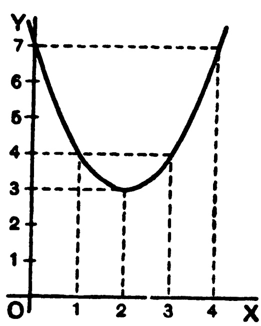

FIG. 26.

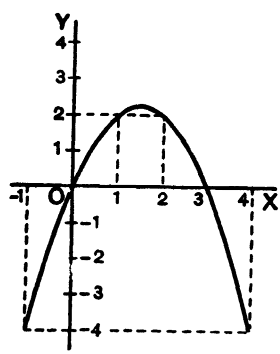

FIG. 27.

minimum height, at that point its $\frac{dy}{dx} = 0$. Now this gives us the clue to the dodge that is wanted. When there is put before you an equation, and you want to find that value of x that will make its y a minimum (or a maximum), *first differentiate it,* and having done so, write its $\frac{dy}{dx}$ as *equal to zero,* and then solve for x. Put this particular value of x into the original equation, and you will then get the required value of y. This process is commonly called "equating to zero".

To see how simply it works, take the example with which this chapter opens, namely,

$$y = x^2 - 4x + 7$$

Differentiating, we get:

$$\frac{dy}{dx} = 2x - 4$$

Now equate this to zero, thus:

$$2x - 4 = 0$$

Solving this equation for x, we get:

$$2x = 4, \quad x = 2$$

Now, we know that the maximum (or minimum) will occur exactly when $x = 2$.

Putting the value $x = 2$ into the original equation, we get

$$y = 2^2 - (4 \times 2) + 7$$
$$= 4 - 8 + 7 = 3$$

Now look back at Fig. 26, and you will see that the minimum occurs when $x = 2$, and that this minimum of $y = 3$.

Try the second example (Fig. 27), which is

$$y = 3x - x^2$$

Differentiating, $\frac{dy}{dx} = 3 - 2x$

Equating to zero, $3 - 2x = 0$

whence $x = 1\frac{1}{2}$

and putting this value of x into the original equation, we find:

$$y = 4\tfrac{1}{2} - \left(1\tfrac{1}{2} \times 1\tfrac{1}{2}\right)$$
$$y = 2\tfrac{1}{4}$$

This gives us exactly the information as to which the method of trying a lot of values left us uncertain.

Now, before we go on to any further cases, we have two remarks to make. When you are told to equate $\frac{dy}{dx}$ to zero, you feel at first (that is if you have any wits of your own) a kind of resentment, because you know that $\frac{dy}{dx}$ has all sorts of different values at different parts of the curve, according to whether it is sloping up or down. So, when you are suddenly told to write

$$\frac{dy}{dx} = 0$$

you resent it, and feel inclined to say that it can't be true. Now you will have to understand the essential difference between "an equation", and "an equation of condition". Ordinarily you are dealing with equations that are true in themselves; but, on occasions, of which the present are examples, you have to write down equations that are not necessarily true, but are only true if certain conditions are to be fulfilled; and you write them down in order, by solving them, to find the conditions which make them true. Now we want to find the particular value that x has when the curve is neither sloping up nor sloping down, that is, at the particular place where $\frac{dy}{dx} = 0$. So, writing $\frac{dy}{dx} = 0$ does *not* mean that it always is $= 0$; but you write it down *as a condition* in order to see how much x will come out if $\frac{dy}{dx}$ is to be zero.

The second remark is one which (if you have any wits of your own) you will probably have already made: namely, that this much-belauded process of equating to zero entirely fails to tell you whether the x that you thereby find is going to give you a *maximum* value of y or a *minimum* value of y. Quite so. It does not of itself discriminate; it finds for you the right value of x but leaves you to find out for yourselves whether the corresponding y is a maximum or a minimum. Of course, if you have plotted the curve, you know already which it will be.

For instance, take the equation:

$$y = 4x + \frac{1}{x}$$

Without stopping to think what curve it corresponds to, differentiate it, and equate to zero:

$$\frac{dy}{dx} = 4 - x^{-2} = 4 - \frac{1}{x^2} = 0$$

whence $$x = \tfrac{1}{2} \quad \text{or} \quad x = -\tfrac{1}{2}$$

and, inserting these values,

$$y = 4 \quad \text{or} \quad y = -4$$

Each will be either a maximum or else a minimum. But which? You will hereafter be told a way, depending upon a second differentiation (see Chap. XII). But at present it is enough if you will simply try two other values of x differing a little from the one found, one larger and one smaller, and see whether with these altered values the corresponding values of y are less or greater than that already found.

Try another simple problem in maxima and minima. Suppose you were asked to divide any number into two parts, such that the product was a maximum? How would you set about it if you did not know the trick of equating to zero? I suppose you could worry it out by the rule of try, try, try again. Let 60 be the number. You can try cutting it into two parts, and multiplying them together. Thus, 50 times 10 is 500; 52 times 8 is 416; 40 times 20 is 800; 45 times 15 is 675; 30 times 30 is 900. This looks like a maximum: try varying it. 31 times 29 is 899, which is not so good; and 32 times 28 is 896, which is worse. So it seems that the biggest product will be got by dividing into two halves.

Now see what the calculus tells you. Let the number to be cut into two parts be called n. Then if x is one part, the other will be $n - x$, and the product will be $x(n - x)$ or $nx - x^2$. So we write $y = nx - x^2$. Now differentiate and equate to zero;

$$\frac{dy}{dx} = n - 2x = 0$$

Solving for x, we get $\quad \dfrac{n}{2} = x$

So now we *know* that whatever number n may be, we must divide it into two equal parts if the product of the parts is to be a maximum; and the value of that maximum product will always be $= \frac{1}{4}n^2$.

This is a very useful rule, and applies to any number of factors, so that if $m + n + p =$ a constant number, $m \times n \times p$ is a maximum when $m = n = p$.

Test Case.

Let us at once apply our knowledge to a case that we can test.

Let $\qquad y = x^2 - x$

and let us find whether this function has a maximum or minimum; and if so, test whether it is a maximum or a minimum.

Differentiating, we get

$$\frac{dy}{dx} = 2x - 1$$

Equating to zero, we get

$$2x - 1 = 0$$

whence $$2x = 1$$

or $$x = \tfrac{1}{2}$$

That is to say, when x is made $= \tfrac{1}{2}$, the corresponding value of y will be either a maximum or a minimum. Accordingly, putting $x = \tfrac{1}{2}$ in the original equation, we get

$$y = \left(\tfrac{1}{2}\right)^2 - \tfrac{1}{2}$$

or $$y = -\tfrac{1}{4}$$

Is this a maximum or a minimum? To test it, try putting x a little bigger than $\tfrac{1}{2}$—say, make $x = 0.6$. Then

$$y = (0.6)^2 - 0.6 = 0.36 - 0.6 = -0.24$$

Also try a value of x a little smaller than $\tfrac{1}{2}$—say, $x = 0.4$.

Then $$y = (0.4)^2 - 0.4 = 0.16 - 0.4 = -0.24$$

Both values are larger than -0.25, showing that $y = -0.25$ is a *minimum.*

Plot the curve for yourself, and verify the calculation.

Further Examples.

A most interesting example is afforded by a curve that has both a maximum and a minimum. Its equation is:

$$y = \tfrac{1}{3}x^3 - 2x^2 + 3x + 1$$

Now $$\frac{dy}{dx} = x^2 - 4x + 3$$

Equating to zero, we get the quadratic,

$$x^2 - 4x + 3 = 0$$

and solving the quadratic gives us *two* roots, viz.

$$\begin{cases} x = 3 \\ x = 1 \end{cases}$$

Now, when $x = 3$, $y = 1$; and when $x = 1$, $y = 2\frac{1}{3}$. The first of these is a minimum, the second a maximum.[1]

The curve itself may be plotted (as in Fig. 28) from the values calculated, as below, from the original equation.

x	-1	0	1	2	3	4	5	6
y	$-4\frac{1}{3}$	1	$2\frac{1}{3}$	$1\frac{2}{3}$	1	$2\frac{1}{3}$	$7\frac{2}{3}$	19

A further exercise in maxima and minima is afforded by the following example:

The equation of a circle of radius r, having its center C at the point whose coordinates are $x = a$, $y = b$, as depicted in Fig. 29, is:

$$(y - b)^2 + (x - a)^2 = r^2$$

This may be transformed into

$$y = \sqrt{r^2 - (x - a)^2} + b$$

(where the square root may be either positive or negative).

1. The maximum point on the curve in Figure 28 is called a "local" maximum because the curve obviously has higher points later on. Similarly, its minimum point is a "local" minimum because there are lower points earlier on the curve. If maximum and minimum points are the highest and lowest points on a curve, as in Figures 26 and 27, they are called "absolute" maxima and minima. In Figure 28 the curve has no absolute maximum or minimum points because it goes to infinity at both ends.

In Figure 28 the point on the curve at $x = 2$ is called an "inflection" point. This is a point at which the curve is concave upward on one side and concave downward on the other. Put another way, it is the point at which a tangent to the line, as you move left to right, stops rotating in one direction and starts to rotate the other way. When a curve's second derivative is positive, the curve is concave upward (like a smile), and when the second derivative is negative, the curve is concave downward (like a frown). Calculus texts often distinguish the two curves by saying that one "holds water" and the other does not.

Sometimes a curve has an inflection point at a point where $\frac{dy}{dx} = 0$. In this case, the point is neither a maximum or minimum. For example, this is true for $y = x^3$ at $x = 0$. So after we equate the derivative to zero, we need to check points on both sides of x to determine if we've found a maximum, a minimum, or neither.—M.G.

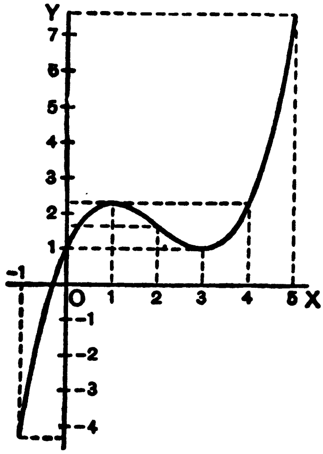

FIG. 28.

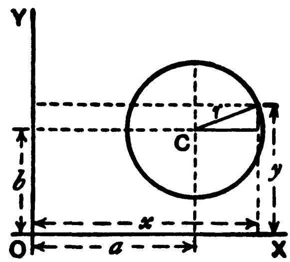

FIG. 29.

Now we know beforehand, by mere inspection of the figure, that when $x = a$, y will be either at its maximum value, $b + r$, or else at its minimum value, $b - r$. But let us not take advantage of this knowledge; let us set about finding what value of x will make y a maximum or a minimum, by the process of differentiating and equating to zero.

$$\frac{dy}{dx} = \frac{1}{2} \frac{1}{\sqrt{r^2 - (x - a)^2}} \times (2a - 2x)$$

which reduces to

$$\frac{dy}{dx} = \frac{a - x}{\sqrt{r^2 - (x - a)^2}}$$

Then the condition for y being maximum or minimum is:

$$\frac{a - x}{\sqrt{r^2 - (x - a)^2}} = 0$$

Since no value whatever of x will make the denominator infinite, the only condition to give zero is

$$x = a$$

Inserting this value in the original equation for the circle, we find

$$y = \sqrt{r^2} + b$$

and as the root of r^2 is either $+r$ or $-r$, we have two resulting values of y,

$$\begin{cases} y = b + r \\ y = b - r \end{cases}$$

The first of these is the maximum, at the top; the second the minimum, at the bottom.

If the curve is such that there is no place that is a maximum or minimum, the process of equating to zero will yield an impossible result. For instance:

Let $$y = ax^3 + bx + c$$

Then $$\frac{dy}{dx} = 3ax^2 + b$$

Equating this to zero, we get $3ax^2 + b = 0$, $x^2 = \frac{-b}{3a}$, and $x = \sqrt{\frac{-b}{3a}}$, which is impossible, supposing a and b to have the same sign.

Therefore y has no maximum nor minimum.

A few more worked examples will enable you to thoroughly master this most interesting and useful application of the calculus.[2]

(1) What are the sides of the rectangle of maximum area inscribed in a circle of radius R?

2. This is an understatement. Finding extrema (maxima and minima) values of a function is one of the most beautiful and useful aspects of differential calculus. Simply equate a derivative to zero and solve for x! It seems to work like magic!—M.G.

If one side be called x,

$$\text{the other side} = \sqrt{(\text{diagonal})^2 - x^2};$$

and as the diagonal of the rectangle is necessarily a diameter of the circumscribing circle, the other side $= \sqrt{4R^2 - x^2}$.

Then, area of rectangle $S = x\sqrt{4R^2 - x^2}$,

$$\frac{dS}{dx} = x \times \frac{d\left(\sqrt{4R^2 - x^2}\right)}{dx} + \sqrt{4R^2 - x^2} \times \frac{d(x)}{dx}$$

If you have forgotten how to differentiate $\sqrt{4R^2 - x^2}$, here is a hint: write $w = 4R^2 - x^2$ and $y = \sqrt{w}$, and seek $\frac{dy}{dw}$ and $\frac{dw}{dx}$; fight it out, and only if you can't get on refer to Chapter IX.

You will get

$$\frac{dS}{dx} = x \times -\frac{x}{\sqrt{4R^2 - x^2}} + \sqrt{4R^2 - x^2} = \frac{4R^2 - 2x^2}{\sqrt{4R^2 - x^2}}$$

For maximum or minimum we must have

$$\frac{4R^2 - 2x^2}{\sqrt{4R^2 - x^2}} = 0$$

that is, $4R^2 - 2x^2 = 0$ and $x = R\sqrt{2}$.

The other side $= \sqrt{4R^2 - 2R^2} = R\sqrt{2}$; the two sides are equal; the figure is a square the side of which is equal to the diagonal of the square constructed on the radius. In this case it is, of course, a maximum with which we are dealing.

(2) What is the radius of the opening of a conical vessel the sloping side of which has a length l when the capacity of the vessel is greatest?

If R be the radius and H the corresponding height,

$$H = \sqrt{l^2 - R^2}$$

$$\text{Volume } V = \pi R^2 \times \frac{H}{3} = \pi R^2 \times \frac{\sqrt{l^2 - R^2}}{3}$$

Proceeding as in the previous problem, we get

$$\frac{dV}{dR} = \pi R^2 \times -\frac{R}{3\sqrt{l^2 - R^2}} + \frac{2\pi R}{3}\sqrt{l^2 - R^2}$$

$$= \frac{2\pi R(l^2 - R^2) - \pi R^3}{3\sqrt{l^2 - R^2}} = 0$$

for maximum or minimum.

Or, $2\pi R(l^2 - R^2) - \pi R^3 = 0$, and $R = l\sqrt{\frac{2}{3}}$, for a maximum, obviously.

(3) Find the maxima and minima of the function

$$y = \frac{x}{4 - x} + \frac{4 - x}{x}$$

We get $$\frac{dy}{dx} = \frac{(4 - x) - (-x)}{(4 - x)^2} + \frac{-x - (4 - x)}{x^2} = 0$$

for maximum or minimum; or

$$\frac{4}{(4 - x)^2} - \frac{4}{x^2} = 0 \quad \text{and} \quad x = 2$$

There is only one value, hence only one maximum or minimum.

For	$x = 2$	$y = 2$
for	$x = 1.5$	$y = 2.27$
for	$x = 2.5$	$y = 2.27$

it is therefore a minimum. (It is instructive to plot the graph of the function.)

(4) Find the maxima and minima of the function

$$y = \sqrt{1 + x} + \sqrt{1 - x}$$

(It will be found instructive to plot the graph.)

Differentiating gives at once (see example No. 1, Chapter IX).

$$\frac{dy}{dx} = \frac{1}{2\sqrt{1 + x}} - \frac{1}{2\sqrt{1 - x}} = 0$$

for maximum or minimum.

Hence $\sqrt{1+x} = \sqrt{1-x}$ and $x = 0$, the only solution

For $x = 0$, $\qquad y = 2.$

For $x = \pm 0.5$, $\quad y = 1.932$, so this is a maximum.

(5) Find the maxima and minima of the function

$$y = \frac{x^2 - 5}{2x - 4}$$

We have $\qquad \dfrac{dy}{dx} = \dfrac{(2x-4) \times 2x - (x^2 - 5)2}{(2x-4)^2} = 0$

for maximum or minimum; or

$$\frac{2x^2 - 8x + 10}{(2x-4)^2} = 0$$

or $x^2 - 4x + 5 = 0$; which has solutions

$$x = 2 \pm \sqrt{-1}$$

These being imaginary, there is no real value of x for which $\dfrac{dy}{dx} = 0$; hence there is neither maximum nor minimum.

(6) Find the maxima and minima of the function

$$(y - x^2)^2 = x^5$$

This may be written $y = x^2 \pm x^{\frac{5}{2}}$.

$$\frac{dy}{dx} = 2x \pm \tfrac{5}{2}x^{\frac{3}{2}} = 0 \text{ for maximum or minimum;}$$

that is, $x\left(2 \pm \frac{5}{2}x^{\frac{1}{2}}\right) = 0$, which is satisfied for $x = 0$, and for $2 \pm \frac{5}{2}x^{\frac{1}{2}} = 0$, that is for $x = \frac{16}{25}$. So there are two solutions.

Taking first $x = 0$. If $x = -0.5$, $y = 0.25 \pm \sqrt[2]{-(.5)^5}$, and if $x = +0.5$, $y = 0.25 \pm \sqrt[2]{(.5)^5}$. On one side y is imaginary; that is, there is no value of y that can be represented by a graph; the latter is therefore entirely on the right side of the axis of y (see Fig. 30).

On plotting the graph it will be found that the curve goes to the origin, as if there were a minimum there; but instead of continuing beyond, as it should do for a minimum, it retraces its steps (forming a cusp). There is no minimum, therefore, although the condition for a minimum is satisfied, namely $\frac{dy}{dx} = 0$. It is necessary therefore always to check by taking one value on either side.[3]

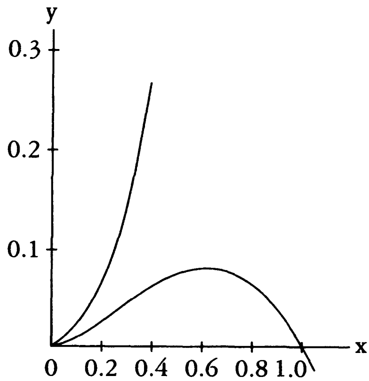

FIG. 30.

Now, if we take $x = \frac{16}{25} = 0.64$. If $x = 0.64$, $y = 0.7373$ and $y = 0.0819$; if $x = 0.6$, y becomes 0.6389 and 0.0811; and if $x = 0.7$, y becomes 0.9000 and 0.0800.

This shows that there are two branches of the curve; the upper one does not pass through a maximum, but the lower one does.

[At this point Thompson introduces a problem that has nothing to do with extrema, but I have let it remain. Note also that problem 10 of the exercises is also out of place for the same reason.—M.G.]

(7) A cylinder whose height is twice the radius of the base is increasing in volume, so that all its parts keep always in the same proportion to each other; that is, at any instant, the cylinder is *similar* to the original cylinder. When the radius of the base is r inches, the surface area is increasing at the rate of 20 square inches per second; at what rate per second is its volume then increasing?

$$\text{Area} = S = 2(\pi r^2) + 2\pi r \times 2r = 6\pi r^2$$

$$\text{Volume} = V = \pi r^2 \times 2r = 2\pi r^3$$

3. Today's terminology broadens the meaning of extrema, allowing local extrema at cusps, corners, and endpoints of intervals in a function's domain.—M.G.

$$\frac{dS}{dt} = 12\pi r\frac{dr}{dt} = 20; \quad \frac{dr}{dt} = \frac{20}{12\pi r}$$

$$\frac{dV}{dt} = 6\pi r^2\frac{dr}{dt}; \quad \text{and}$$

$$\frac{dV}{dt} = 6\pi r^2 \times \frac{20}{12\pi r} = 10r$$

The volume changes at the rate of $10r$ cubic inches per second.

Make other examples for yourself. There are few subjects which offer such a wealth for interesting examples.

EXERCISES IX

(1) What values of x will make y a maximum and a minimum, if $y = \dfrac{x^2}{x+1}$?

(2) What value of x will make y a maximum in the equation $y = \dfrac{x}{a^2+x^2}$?

(3) A line of length p is to be cut up into 4 parts and put together as a rectangle. Show that the area of the rectangle will be a maximum if each of its sides is equal to $\frac{1}{4}p$.

(4) A piece of string 30 inches long has its two ends joined together and is stretched by 3 pegs so as to form a triangle. What is the largest triangular area that can be enclosed by the string?

(5) Plot the curve corresponding to the equation

$$y = \frac{10}{x} + \frac{10}{8-x}$$

also find $\dfrac{dy}{dx}$, and deduce the value of x that will make y a minimum; and find that minimum value of y.

(6) If $y = x^5 - 5x$, find what values of x will make y a maximum or a minimum.

(7) What is the smallest square that can be placed in a given square so each corner of the small square touches a side of the larger square?

(8) Inscribe in a given cone, the height of which is equal to the radius of the base, a cylinder (*a*) whose volume is a maximum; (*b*) whose lateral area is a maximum; (*c*) whose total area is a maximum.

(9) Inscribe in a sphere, a cylinder (*a*) whose volume is a maximum; (*b*) whose lateral area is a maximum; (*c*) whose total area is a maximum.

(10) A spherical balloon is increasing in volume. If, when its radius is r feet, its volume is increasing at the rate of 4 cubic feet per second, at what rate is its surface then increasing?

(11) Inscribe in a given sphere a cone whose volume is a maximum.

Chapter XII

CURVATURE OF CURVES

Returning to the process of successive differentiation, it may be asked: Why does anybody want to differentiate twice over? We know that when the variable quantities are space and time, by differentiating twice over we get the acceleration of a moving body, and that in the geometrical interpretation, as applied to curves, $\dfrac{dy}{dx}$ means the *slope* of the curve. But what can $\dfrac{d^2y}{dx^2}$ mean in this case? Clearly it means the rate (per unit of length x) at which the slope is changing—in brief, it is *an indication of the manner in which the slope of the portion of curve considered varies,* that is, whether the slope of the curve increases or decreases when x increases, or, in other words, whether the curve curves up or down towards the right.

Suppose a slope constant, as in Fig. 31.

Here, $\dfrac{dy}{dx}$ is of constant value.

Suppose, however, a case in which, like Fig. 32, the slope itself

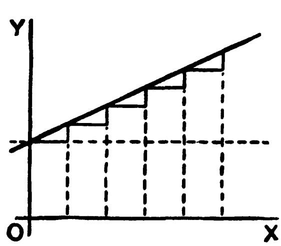

FIG. 31.

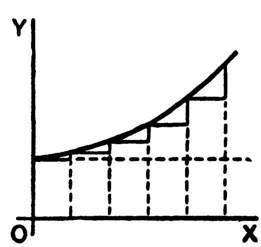

FIG. 32.

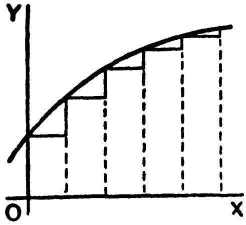

FIG. 33.

is getting greater upwards; then $\dfrac{d\left(\dfrac{dy}{dx}\right)}{dx}$, that is, $\dfrac{d^2y}{dx^2}$, will be *positive.*

If the slope is becoming less as you go to the right (as in Fig. 14), or as in Fig. 33, then, even though the curve may be going upward, since the change is such as to diminish its slope, its $\dfrac{d^2y}{dx^2}$ will be *negative.*

It is now time to initiate you into another secret—how to tell whether the result that you get by "equating to zero" is a maximum or a minimum. The trick is this: After you have differentiated (so as to get the expression which you equate to zero), you then differentiate a second time and look whether the result of the second differentiation is *positive* or *negative.* If $\dfrac{d^2y}{dx^2}$ comes out *positive,* then you know that the value of y which you got was a *minimum*; but if $\dfrac{d^2y}{dx^2}$ comes out *negative,* then the value of y which you got must be a *maximum.* That's the rule.

The reason of it ought to be quite evident. Think of any curve that has a minimum point in it, like Fig. 15, or like Fig. 34, where the point of minimum y is marked M, and the curve is *concave* upward. To the left of M the slope is downward, that is, negative, and is getting less negative. To the right of M the slope has become upward, and is getting more and more upward. Clearly the change of slope as the curve passes through M is such that $\dfrac{d^2y}{dx^2}$ is *positive,* for its operation, as x increases toward the right, is to convert a downward slope into an upward one.

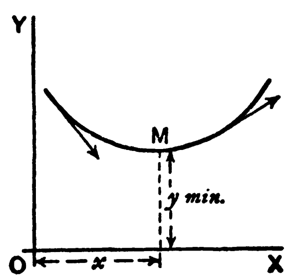

FIG. 34.

Similarly, consider any curve that has a maximum point in it, like Fig. 16, or like Fig. 35, where the curve is *concave upward,* and the maximum point is marked M. In this case, as the curve passes through M from left to right, its upward slope is converted into a downward or negative slope, so that in this case the "slope of the slope" $\frac{d^2y}{dx^2}$ is *negative.*

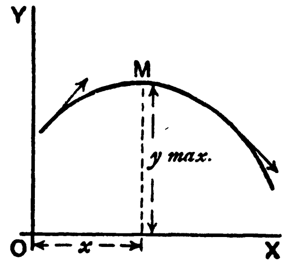

FIG. 35.

Go back now to the examples of the last chapter and verify in this way the conclusions arrived at as to whether in any particular case there is a maximum or a minimum. You will find below a few worked-out examples.

(1) Find the maximum or minimum of

$$(a)\ y = 4x^2 - 9x - 6; \quad (b)\ y = 6 + 9x - 4x^2$$

and ascertain if it be a maximum or a minimum in each case.

$$(a)\quad \frac{dy}{dx} = 8x - 9 = 0;\ x = 1\tfrac{1}{8};\ \text{and } y = -11.0625$$

$$\frac{d^2y}{dx^2} = 8;\ \text{it is } +;\ \text{hence it is a minimum.}$$

$$(b)\quad \frac{dy}{dx} = 9 - 8x = 0;\ x = 1\tfrac{1}{8};\ \text{and } y = +11.0625$$

$$\frac{d^2y}{dx^2} = -8;\ \text{it is } -;\ \text{hence it is a maximum.}$$

(2) Find the maxima and minima of the function

$$y = x^3 - 3x + 16$$

$$\frac{dy}{dx} = 3x^2 - 3 = 0;\ x^2 = 1;\ \text{and } x = \pm 1$$

$$\frac{d^2y}{dx^2} = 6x; \text{ for } x = 1; \text{ it is } +,$$

hence $x = 1$ corresponds to a minimum $y = 14$. For $x = -1$ it is $-$; hence $x = -1$ corresponds to a maximum $y = +18$.

(3) Find the maxima and minima of $y = \dfrac{x-1}{x^2+2}$.

$$\frac{dy}{dx} = \frac{(x^2+2)\times 1 - (x-1)\times 2x}{(x^2+2)^2} = \frac{2x - x^2 + 2}{(x^2+2)^2} = 0$$

or $x^2 - 2x - 2 = 0$, whose solutions are $x = +2.73$ and $x = -0.73$.

$$\frac{d^2y}{dx^2} = -\frac{(x^2+2)^2(2-2x) - (2x - x^2 + 2)(4x^3 + 8x)}{(x^2+2)^4}$$

$$= \frac{2x^5 - 6x^4 - 8x^3 - 8x^2 - 24x + 8}{(x^2+2)^4}$$

The denominator is always positive, so it is sufficient to ascertain the sign of the numerator.

If we put $x = 2.73$, the numerator is negative; the maximum, $y = 0.183$.

If we put $x = -0.73$, the numerator is positive; the minimum, $y = -0.683$.

(4) The expense C of handling the products of a certain factory varies with the weekly output P according to the relation $C = aP + \dfrac{b}{c+P} + d$, where a, b, c, d are positive constants. For what output will the expense be least?

$\dfrac{dC}{dP} = a - \dfrac{b}{(c+P)^2} = 0$ for maximum or minimum; hence $a = \dfrac{b}{(c+P)^2}$ and $P = \pm\sqrt{\dfrac{b}{a}} - c$

As the output cannot be negative, $P = +\sqrt{\dfrac{b}{a}} - c$.

Now
$$\frac{d^2C}{dP^2} = +\frac{b(2c+2P)}{(c+P)^4}$$

which is positive for all the values of P; hence $P = +\sqrt{\frac{b}{a}} - c$ corresponds to a minimum.

(5) The total cost per hour C of lighting a building with N lamps of a certain kind is

$$C = N\left(\frac{C_l}{t} + \frac{EPC_e}{1000}\right)$$

where E is the commercial efficiency (watts per candle),

P is the candle power of each lamp,

t is the average life of each lamp in hours,

$C_l =$ cost of renewal in cents per hour of use,

$C_e =$ cost of energy per 1000 watts per hour.

Moreover, the relation connecting the average life of a lamp with the commercial efficiency at which it is run is approximately $t = mE^n$, where m and n are constants depending on the kind of lamp.

Find the commercial efficiency for which the total cost of lighting will be least.

We have
$$C = N\left(\frac{C_l}{m}E^{-n} + \frac{PC_e}{1000}E\right)$$

$$\frac{dC}{dE} = N\left(\frac{PC_e}{1000} - \frac{nC_l}{m}E^{-(n+1)}\right) = 0$$

for maximum or minimum.

$$E^{n+1} = \frac{1000 \times nC_l}{mPC_e} \quad \text{and} \quad E = \sqrt[n+1]{\frac{1000 \times nC_l}{mPC_e}}$$

This is clearly for minimum, since

$$\frac{d^2C}{dE^2} = N\left[(n+1)\frac{nC_l}{m}E^{-(n+2)}\right]$$

which is positive for a positive value of E.

For a particular type of 16 candle-power lamps, $C_l = 17$ cents, $C_e = 5$ cents; and it was found that $m = 10$ and $n = 3.6$.

$$E = \sqrt[4.6]{\frac{1000 \times 3.6 \times 17}{10 \times 16 \times 5}} = 2.6 \text{ watts per candle power.}$$

EXERCISES X

You are advised to plot the graph of any numerical example.

(1) Find the maxima and minima of

$$y = x^3 + x^2 - 10x + 8$$

(2) Given $y = \frac{b}{a}x - cx^2$, find expressions for $\frac{dy}{dx}$, and for $\frac{d^2y}{dx^2}$; also find the value of x which makes y a maximum or a minimum, and show whether it is maximum or minimum. Assume $c > 0$.

(3) Find how many maxima and how many minima there are in the curve, the equation to which is

$$y = 1 - \frac{x^2}{2} + \frac{x^4}{24}$$

and how many in that of which the equation is

$$y = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720}$$

(4) Find the maxima and minima of

$$y = 2x + 1 + \frac{5}{x^2}$$

(5) Find the maxima and minima of

$$y = \frac{3}{x^2 + x + 1}$$

(6) Find the maxima and minima of

$$y = \frac{5x}{2 + x^2}$$

(7) Find the maxima and minima of

$$y = \frac{3x}{x^2 - 3} + \frac{x}{2} + 5$$

(8) Divide a number N into two parts in such a way that three times the square of one part plus twice the square of the other part shall be a minimum.

(9) The efficiency u of an electric generator at different values of output x is expressed by the general equation:

$$u = \frac{x}{a + bx + cx^2}$$

where a is a constant depending chiefly on the energy losses in the iron and c a constant depending chiefly on the resistance of the copper parts. Find an expression for that value of the output at which the efficiency will be a maximum.

(10) Suppose it to be known that consumption of coal by a certain steamer may be represented by the formula

$$y = 0.3 + 0.001v^3$$

where y is the number of tons of coal burned per hour and v is the speed expressed in nautical miles per hour. The cost of wages, interest on capital, and depreciation of that ship are together equal, per hour, to the cost of 1 ton of coal. What speed will make the total cost of a voyage of 1000 nautical miles a minimum? And, if coal costs 10 dollars per ton, what will that minimum cost of the voyage amount to?

(11) Find the maxima and minima of

$$y = \pm\frac{x}{6}\sqrt{x(10 - x)}$$

(12) Find the maxima and minima of

$$y = 4x^3 - x^2 - 2x + 1$$

Chapter XIII

PARTIAL FRACTIONS AND INVERSE FUNCTIONS

We have seen that when we differentiate a fraction we have to perform a rather complicated operation; and, if the fraction is not itself a simple one, the result is bound to be a complicated expression. If we could split the fraction into two or more simpler fractions such that their sum is equivalent to the original fraction, we could then proceed by differentiating each of these simpler expressions. And the result of differentiating would be the sum of two (or more) derivatives, each one of which is relatively simple; while the final expression, though of course it will be the same as that which could be obtained without resorting to this dodge, is thus obtained with much less effort and appears in a simplified form.

Let us see how to reach this result. Try first the job of adding two fractions together to form a resultant fraction. Take, for example, the two fractions $\dfrac{1}{x+1}$ and $\dfrac{2}{x-1}$. Every schoolboy can add these together and find their sum to be $\dfrac{3x+1}{x^2-1}$. And in the same way he can add together three or more fractions. Now this process can certainly be reversed: that is to say that, if this last expression were given, it is certain that it can somehow be split back again into its original components or partial fractions. Only we do not know in every case that may be presented to us *how* we can so split it. In order to find this out we shall consider a simple case at first. But it is important to bear in mind that all which follows applies only to what are called "proper" algebraic fractions, meaning fractions like the above, which have the numerator of *a lesser degree* than the denominator; that is, those in which the highest exponent of x is less in the numerator than in the denominator. If we have to deal with such an expression as $\dfrac{x^2+2}{x^2-1}$, we can simplify it by division, since it is equivalent to $1+\dfrac{3}{x^2-1}$; and $\dfrac{3}{x^2-1}$ is a proper algebraic fraction to which the operation of splitting into partial fractions can be applied, as explained hereafter.

Case I. If we perform many additions of two or more fractions the denominators of which contain only terms in x, and no terms in x^2, x^3, or any other powers of x, we *always* find that *the denominator of the final resulting fraction is the product of the denominators* of the fractions which were added to form the result. It follows that by factorizing the denominator of this final fraction, we can find every one of the denominators of the partial fractions of which we are in search.

Suppose we wish to go back from $\dfrac{3x+1}{x^2-1}$ to the components which we know are $\dfrac{1}{x+1}$ and $\dfrac{2}{x-1}$. If we did not know what those components were we can still prepare the way by writing:

$$\frac{3x+1}{x^2-1}=\frac{3x+1}{(x+1)(x-1)}=\frac{\quad}{x+1}+\frac{\quad}{x-1}$$

leaving blank the places for the numerators until we know what to put there. We always may assume the sign between the partial fractions to be *plus,* since, if it be *minus,* we shall simply find the corresponding numerator to be negative. Now, since the partial fractions are *proper* fractions, the numerators are mere numbers without x at all, and we can call them A, B, C . . . as we please. So, in this case, we have:

$$\frac{3x+1}{x^2-1}=\frac{A}{x+1}+\frac{B}{x-1}$$

If, now, we perform the addition of these two partial fractions, we get $\dfrac{A(x-1)+B(x+1)}{(x+1)(x-1)}$; and this must be equal to $\dfrac{3x+1}{(x+1)(x-1)}$. And, as the denominators in these two expressions are the same, the numerators must be equal, giving us:

$$3x+1=A(x-1)+B(x+1)$$

Now, this is an equation with two unknown quantities, and it would seem that we need another equation before we can solve them and find A and B. But there is another way out of this difficulty. The equation must be true for all values of x; therefore it must be true for such values of x as will cause $x-1$ and $x+1$ to become zero, that is for $x=1$ and for $x=-1$ respectively. If we make $x=1$, we get $4=(A\times 0)+(B\times 2)$, so that $B=2$; and if we make $x=-1$, we get

$$-2=(A\times -2)+(B\times 0)$$

so that $A=1$. Replacing the A and B of the partial fractions by these new values, we find them to become $\dfrac{1}{x+1}$ and $\dfrac{2}{x-1}$; and the thing is done.

As a further example, let us take the fraction

$$\frac{4x^2+2x-14}{x^3+3x^2-x-3}$$

The denominator becomes zero when x is given the value 1; hence $x-1$ is a factor of it, and obviously then the other factor will be x^2+4x+3; and this can again be decomposed into $(x+1)(x+3)$. So we may write the fraction thus:

$$\frac{4x^2+2x-14}{x^3+3x^2-x-3}=\frac{A}{x+1}+\frac{B}{x-1}+\frac{C}{x+3}$$

making three partial factors.

Proceeding as before, we find

$$\begin{aligned}4x^2+2x-14=A(x-1)(x+3)&+B(x+1)(x+3)\\&+C(x+1)(x-1)\end{aligned}$$

Now, if we make $x = 1$, we get:

$$-8 = (A \times 0) + B(2 \times 4) + (C \times 0); \text{ that is, } B = -1$$

If $x = -1$, we get

$$-12 = A(-2 \times 2) + (B \times 0) + (C \times 0); \text{ whence } A = 3.$$

If $x = -3$, we get:

$$16 = (A \times 0) + (B \times 0) + C(-2 \times -4); \text{ whence } C = 2.$$

So then the partial fractions are:

$$\frac{3}{x+1} - \frac{1}{x-1} + \frac{2}{x+3}$$

which is far easier to differentiate with respect to x than the complicated expression from which it is derived.

Case II. If some of the factors of the denominator contain terms in x^2, and are not conveniently put into factors, then the corresponding numerator may contain a term in x, as well as a simple number, and hence it becomes necessary to represent this unknown numerator not by the symbol A but by $Ax + B$; the rest of the calculation being made as before.

Try, for instance: $\dfrac{-x^2 - 3}{(x^2 + 1)(x + 1)}$

$$\frac{-x^2 - 3}{(x^2 + 1)(x + 1)} = \frac{Ax + B}{x^2 + 1} + \frac{C}{x + 1}$$

$$-x^2 - 3 = (Ax + B)(x + 1) + C(x^2 + 1)$$

Putting $x = -1$, we get $-4 = C \times 2$; and $C = -2$

hence $\quad -x^2 - 3 = (Ax + B)(x + 1) - 2x^2 - 2$

and $\quad x^2 - 1 = Ax(x + 1) + B(x + 1)$

Putting $x = 0$, we get $-1 = B$;

hence $\quad x^2 - 1 = Ax(x + 1) - x - 1; \quad$ or $\quad x^2 + x = Ax(x + 1)$

and $\quad x + 1 = A(x + 1)$

so that $A = 1$, and the partial fractions are:

$$\frac{x-1}{x^2+1} - \frac{2}{x+1}$$

Take as another example the fraction

$$\frac{x^3-2}{(x^2+1)(x^2+2)}$$

We get $$\frac{x^3-2}{(x^2+1)(x^2+2)} = \frac{Ax+B}{x^2+1} + \frac{Cx+D}{x^2+2}$$

$$= \frac{(Ax+B)(x^2+2)+(Cx+D)(x^2+1)}{(x^2+1)(x^2+2)}$$

In this case the determination of A, B, C, D is not so easy. It will be simpler to proceed as follows: Since the given fraction and the fraction found by adding the partial fractions are equal, and have *identical* denominators, the numerators must also be identically the same. In such a case, and for such algebraical expressions as those with which we are dealing here, *the coefficients of the same powers of x are equal and of same sign.*

Hence, since

$$x^3 - 2 = (Ax+B)(x^2+2) + (Cx+D)(x^2+1)$$
$$= (A+C)x^3 + (B+D)x^2 + (2A+C)x + 2B + D$$

we have $1 = A + C$; $0 = B + D$ (the coefficient of x^2 in the left expression being zero); $0 = 2A + C$; and $-2 = 2B + D$. Here are four equations, from which we readily obtain $A = -1$; $B = -2$; $C = 2$; $D = 2$; so that the partial fractions are $\frac{2(x+1)}{x^2+2} - \frac{x+2}{x^2+1}$. This method can always be used; but the method shown first will be found the quickest in the case of factors in x only.

Case III. When among the factors of the denominator there are some which are raised to some power, one must allow for the possible existence of partial fractions having for denominator the