1

Introduction

Since antiquity the intuitive notions of continuous change, growth, and motion, have challenged scientific minds. Yet, the way to the understanding of continuous variation was opened only in the seventeenth century when modern science emerged and rapidly developed in close conjunction with integral and differential calculus, briefly called calculus, and mathematical analysis.

The basic notions of Calculus are derivative and integral: the derivative is a measure for the rate of change, the integral a measure for the total effect of a process of continuous change. A precise understanding of these concepts and their overwhelming fruitfulness rests upon the concepts of limit and of function which in turn depend upon an understanding of the continuum of numbers. Only gradually, by penetrating more and more into the substance of Calculus, can one appreciate its power and beauty. In this introductory chapter we shall explain the basic concepts of number, function, and limit, at first simply and intuitively, and then with careful argument.

1.1 The Continuum of Numbers

The positive integers or *natural numbers* 1, 2, 3, . . . are abstract symbols for indicating "how many" objects there are in a *collection* or *set* of discrete elements.

These symbols are stripped of all reference to the concrete qualities of the objects counted, whether they are persons, atoms, houses, or any objects whatever.

The natural numbers are the adequate instrument for counting elements of a collection or "set." However, they do not suffice for another equally important objective: to *measure* quantities such as the length of a curve and the volume or weight of a body. The question, "how much?", cannot be answered immediately in terms of the natural numbers. The profound need for expressing measures of quantities in terms of what we would like to call numbers forces us to extend the number concept so that we may describe a continuous gradation of measures. This extension is called the *number continuum* or the system of "real numbers" (a nondescriptive but generally accepted name). The extension of the number concept to that of the continuum is so convincingly natural that it was used by all the great mathematicians and scientists of earlier times without probing questions. Not until the nineteenth century did mathematicians feel compelled to seek a firmer logical foundation for the real number system. The ensuing precise formulation of the concepts, in turn, led to further progress in mathematics. We shall begin with an unencumbered intuitive approach, and later on we shall give a deeper analysis of the system of real numbers.[1]

a. The System of Natural Numbers and Its Extension. Counting and Measuring

The Natural and the Rational Numbers. The sequence of "natural" numbers 1, 2, 3, . . . is considered as given to us. We need not discuss how these abstract entities, the numbers, may be categorized from a philosophical point of view. For the mathematician, and for anybody working with numbers, it is important merely to know the rules or laws by which they may be combined to yield other natural numbers. These laws form the basis of the familiar rules for adding and multiplying numbers in the decimal system; they include the *commutative laws* $a + b = b + a$ and $ab = ba$, the *associative laws* $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$, the *distributive law* $a(b + c) = ab + ac$, the cancellation law that $a + c = b + c$ implies $a = b$, etc.

The inverse operations, subtraction and division, are not always possible within the set of natural numbers; we cannot subtract 2 from 1 or divide 1 by 2 and stay within that set. To make these operations possible without restriction we are forced to extend the concept of number by inventing the number 0, the "negative" integers, and the fractions. The totality of all these numbers is called the class or set of *rational numbers*; they are all obtained from unity by using the "rational operations" of calculation, namely, addition, subtraction, multiplication, and division.[2]

A rational number can always be written in the form p/q, where p

[1] A more complete exposition is given in *What Is Mathematics?* by Courant and Robbins, Oxford University Press, 1962.

[2] The word "rational" here does not mean reasonable or logical but is derived from the word "ratio" meaning the relative proportion of two magnitudes.

and q are integers and $q \neq 0$. We can make this representation unique by requiring that q is positive and that p and q have no common factor larger than 1.

Within the domain of rational numbers all the *rational operations,* addition, multiplication, subtraction, and division (except division by zero), can be performed and produce again rational numbers. As we know from elementary arithmetic, operations with rational numbers obey the same laws as operations with natural numbers: thus the rational numbers extend the system of positive integers in a completely straightforward way.

Graphical Representation of Rational Numbers. Rational numbers are usually represented graphically by points on a straight line L, the *number axis.* Taking an arbitrary point of L as the origin or point 0

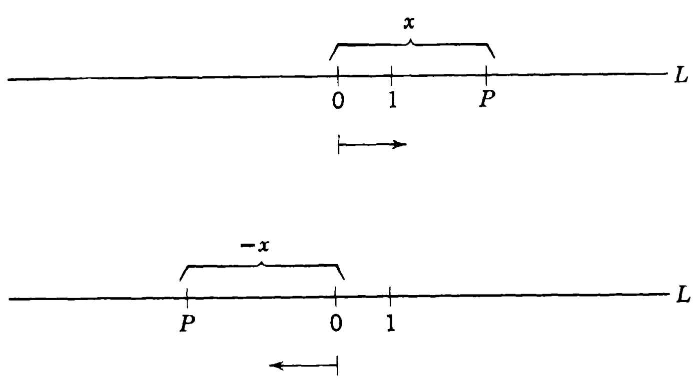

Figure 1.1 The number axis.

and another arbitrary point as the point 1, we use the distance between these two points to serve as a scale or unit of measurement and define the direction from 0 to 1 as "positive." The line with a direction thus imposed is called a directed line. It is customary to depict L so that the point 1 is to the right of the point 0 (Fig. 1.1). The location of any point P on L is completely determined by two pieces of information: the distance of P from the origin 0 and the direction from 0 to P (to the right or left of 0). The point P on L representing a positive rational number lies at distance x units to the right of 0. A negative rational number x is represented by the point $-x$ units to the left of 0. In either case the distance from 0 to the point which represents x is called the absolute value of x, written $|x|$, and we have

$$|x| = \begin{cases} x, & \text{if } x \text{ is positive or zero,} \\ -x, & \text{if } x \text{ is negative.} \end{cases}$$

We note that $|x|$ is never negative and equals zero only when $x = 0$.

From elementary geometry we recall that with ruler and compass it is possible to construct a subdivision of the unit length into any number of equal parts. It follows that any rational length can be constructed and hence that the point representing a rational number x can be found by purely geometrical methods.

In this way we obtain a geometrical representation of rational numbers by points on L, the *rational points.* Consistent with our notation for the points 0 and 1, we take the liberty of denoting both the rational number and the corresponding point on L by the same symbol x.

The relation $x < y$ for two rational numbers means geometrically that the point x lies to the left of the point y. In that case the distance between the points is $y - x$ units. If $x > y$, the distance is $x - y$ units. In either case the distance between two rational points x, y of L is $|y - x|$ units and is again a rational number.

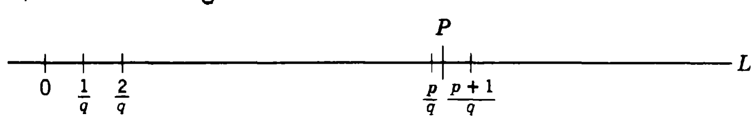

Figure 1.2

A segment on L with end points a, b where $a < b$ will be called an *interval.* The particular segment with end points 0, 1 is called the *unit interval.* If the end points are included in the interval, we say the interval is *closed*; if the end points are excluded, the interval is called *open.* The open interval, denoted by (a, b), consists of those points x for which $a < x < b$, that is, of those points that lie "between" a and b. The closed interval, denoted by $[a, b]$, consists of the points x for which $a \leq x \leq b$.[1] In either case the *length of the interval* is $b - a$.

The points corresponding to the integers $0, \pm 1, \pm 2, \ldots$ subdivide the number axis into intervals of unit length. Every point on L is either an end point or interior point of one of the intervals of the subdivision. If we further subdivide every interval into q equal parts, we obtain a subdivision of L into intervals of length $1/q$ by rational points of the form p/q. Every point P of L is then either a rational point of the form p/q or lies between two successive rational points p/q and $(p + 1)/q$ (see Fig. 1.2). Since successive points of subdivision are $1/q$ units apart, it follows that we can find a rational point p/q whose distance from P does not exceed $1/q$ units. The number $1/q$ can be made as small as we please by choosing q as a sufficiently large positive integer. For example, choosing $q = 10^n$ (where n is any natural number) we can

[1] The relation $a \leq x$ (read "a less than or equal to x") is interpreted as "either $a < x$, or $a = x$." We interpret the double signs $\geq$ and $\pm$ in similar fashion.

find a "decimal fraction" $x = p/10^n$ whose distance from P is less than $1/10^n$. Although we do not assert that every point of L is a rational point we see at least that rational points can be found arbitrarily close to any point P of L.

Density

The arbitrary closeness of rational points to a given point P of L is expressed by saying: *The rational points are dense on the number axis.* It is clear that even smaller sets of rational numbers are dense, for example, the points $x = p/10^n$, for all natural numbers n and integers p.

Density implies that between any two distinct rational points a and b there are infinitely many other rational points. In particular, the point halfway between a and b, $c = \frac{1}{2}(a + b)$, corresponding to the arithmetic mean of the numbers a and b, is again rational. Taking the midpoints of a and c, of b and c, and continuing in this manner, we can obtain any number of rational points between a and b.

An arbitrary point P on L can be located to any degree of precision by using rational points. At first glance it might then seem that the task of locating P by a number has been achieved by introducing the rational numbers. After all, in physical reality quantities are never given or known with absolute precision but always only with a degree of uncertainty and therefore might just as well be considered as measured by rational numbers.

Incommensurable Quantities. Dense as the rational numbers are, they do not suffice as a theoretical basis of measurement by numbers. Two quantities whose ratio is a rational number are called *commensurable* because they can be expressed as integral multiples of a common unit. As early as in the fifth or sixth century B.C. Greek mathematicians and philosophers made the surprising and profoundly exciting discovery that there exist quantities which are not commensurable with a given unit. In particular, line segments exist which are not rational multiples of a given unit segment.

It is easy to give an example of a length incommensurable with the unit length: the diagonal l of a square with the sides of unit length. For, by the theorem of Pythagoras, the square of this length l must be equal to 2. Therefore, if l were a rational number and consequently equal to p/q, where p and q are positive integers, we should have $p^2 = 2q^2$. We can assume that p and q have no common factors, for such common factors could be canceled out to begin with. According to the above equation, p^2 is an even number; hence p itself must be even, say $p = 2p'$. Substituting $2p'$ for p gives us $4p'^2 = 2q^2$, or $q^2 = 2p'^2$; consequently, q^2

is even and so q is also even. This proves that p and q both have the factor 2. However, this contradicts our hypothesis that p and q have no common factor. Since the assumption that the diagonal can be represented by a fraction p/q leads to a contradiction, it is false.

This reasoning, a characteristic example of *indirect proof*, shows that the symbol $\sqrt{2}$ cannot correspond to any rational number. Another example is π, the ratio of the circumference of a circle to its diameter. The proof that π is not rational is much more complicated and was obtained only in modern times (Lambert, 1761). It is easy to find many incommensurable quantities (see Problem 1, p. 106); in fact, incommensurable quantities are in a sense far more common than the commensurable ones (see p. 99).

Irrational Numbers

Because the system of rational numbers is not sufficient for geometry, it is necessary to invent new numbers as measures of incommensurable quantities: these new numbers are called "irrational." The ancient Greeks did not emphasize the abstract number concept, but considered geometric entities, such as line segments, as the basic elements. In a purely geometrical way, they developed a logical system for dealing and operating with incommensurable quantities as well as commensurable (rational) ones. This important achievement, initiated by the Pythagoreans, was greatly advanced by Eudoxus and is expressed at length in Euclid's famous *Elements*. In modern times mathematics was recreated and vastly expanded on a foundation of number concepts rather than geometrical ones. With the introduction of analytic geometry a reversal of emphasis developed in the ancient relationship between numbers and geometrical quantities and the classical theory of incommensurables was all but forgotten or disregarded. It was assumed as a matter of course that to every point on the number axis there corresponds a rational or irrational number and that this totality of "real" numbers obeys the same arithmetical laws as the rational numbers do. Only later, in the nineteenth century, was the need for justifying such an assumption felt and was eventually completely satisfied in a remarkable booklet by Dedekind which makes fascinating reading even today.[1]

---

[1] R. Dedekind, "Nature and Meaning of Number" in *Essays on Number*, London and Chicago, 1901. (The first of these essays, "Continuity and Irrational Numbers," supplies a detailed account of the definition and laws of operation with real numbers.) Reprinted under title *Essays on the Theory of Numbers*, Dover, New York, 1964. The original of these translations appeared in 1887 under the title "Was sind und wass sollen die Zahlen?"

In effect, Dedekind showed that the "naive" approach practiced by all the great mathematicians from Fermat and Newton to Gauss and Riemann was on the right track: That the system of real numbers (as symbols for the lengths of segments, or otherwise defined) is a consistent and complete instrument for scientific measurement, and that in this system the rules of computation of the rational number system remain valid.

Without harm, one could leave it at that and turn directly to the substance of calculus. However, for a deeper understanding of the concept of real number, which is necessary for our later work, the following account as well as the Supplement to this chapter should be studied.

b. Real Numbers and Nested Intervals

For the moment let us think of the points on a line L as the basic elements of the continuum. We postulate that to each point on L there corresponds a "real number" x, its *coordinate*, and that for these numbers x, y the relationships just described for the rational numbers retain their meaning. In particular, the relationship $x < y$ indicates order on L and the expression $|y - x|$ means the distance between the point x and the point y. The basic problem is to relate these numbers (or measurements on the geometrically given continuum of points) to the rational numbers considered originally and hence ultimately to the integers. In addition, we have to explain how to operate with the elements of this "number-continuum" in the same way as with the rational numbers. Eventually, we shall formulate the concept of the continuum of numbers independently of the intuitive geometric concepts, but for the present we postpone some of the more abstract discussion to the Supplement.

How can we describe an irrational real number? For some numbers such as $\sqrt{2}$ or π, we can give a simple geometric characterization, but that is not always feasible. A method flexible enough to yield every real point consists in describing the value x by a sequence of rational approximations of greater and greater precision. Specifically, we shall approximate x simultaneously from the right and from the left with successively increasing accuracy and in such a way that the margin of error approaches zero. In other words, we use a "sequence" of rational intervals containing x, with each interval of the sequence containing the next one, such that the length of the interval, and with it the error of the approximation, can be made smaller than any specified positive number by taking intervals sufficiently far along in the sequence.

To begin, let x be confined to a closed interval $I_1 = [a_1, b_1]$, that is,

$$a_1 \leq x \leq b_1,$$

where a_1 and b_1 are rational (see Fig. 1.3). Within I_1 we consider a "subinterval" $I_2 = [a_2, b_2]$ containing x, that is,

$$a_1 \leq a_2 \leq x \leq b_2 \leq b_1,$$

where a_2 and b_2 are rational. For example, we may choose for I_2 one of the halves of I_1, for x must lie in one or both of the half-intervals. Within I_2 we consider a subinterval $I_3 = [a_3, b_3]$ which also contains x:

$$a_1 \leq a_2 \leq a_3 \leq x \leq b_3 \leq b_2 \leq b_1,$$

where a_3 and b_3 are rational, etc. We require that the length of the interval I_n *tends to zero* with increasing n; that is, that the length of I_n is less than any preassigned positive number for all sufficiently large n. A set of closed intervals $I_1, I_2, I_3, \ldots$ each containing the

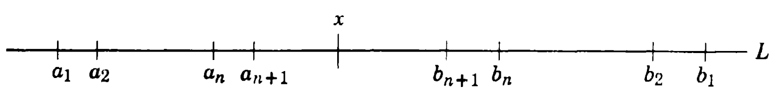

Figure 1.3 A nested sequence of intervals.

next one and such that the lengths tend to zero will be called a "nested sequence of intervals." The point x is uniquely determined by the nested sequence; that is, no other point y can lie in all I_n, since the distance between x and y would exceed the length of I_n once n is sufficiently large. Since here we always choose rational points for the end points of the I_n and since every interval with rational end points is described by two rational numbers, we see that every point x of L, that is, every real number, can be precisely described with the help of infinitely many rational numbers. The converse statement is not so obvious; we shall accept it as a basic *axiom*.

POSTULATE OF NESTED INTERVALS. *If I_1, I_2, $I_3, \ldots$ form a nested sequence of intervals with rational end points, there is a point x contained in all I_n.*[1]

As we shall see, this is an *axiom of continuity*: it guarantees that no gaps exist on the real axis. We shall use the axiom to characterize the real continuum and to justify all operations with limits which are

[1] It is important to emphasize for a nested sequence that the intervals I_n are *closed*. If, for example, I_n denotes the open interval $0 < x < 1/n$, then each I_n contains the following one and the lengths of the intervals tend to zero; but there is no x contained in all I_n.

basic for calculus and analysis. (There also are many other ways of formulating this axiom as we shall see later.)

c. Decimal Fractions. Bases Other than Ten

Infinite Decimal Fractions. One of the many ways of defining real numbers is the familiar description in terms of infinite decimals. It is entirely possible to take the infinite decimals as the basic objects rather than the points of the number axis, but we would rather proceed in a more suggestive geometrical way by defining the infinite decimal representation of real numbers in terms of nested sequences of intervals.

Let the number axis be subdivided into unit intervals by the points corresponding to integers. A point x either lies between two successive points of subdivision or is itself one of the dividing points. In either case there is at least one integer c_0 such that

$$c_0 \leq x \leq c_0 + 1,$$

so that x belongs to the closed interval $I_0 = [c_0, c_0 + 1]$. We divide I_0 into ten equal parts by points $c_0 + \frac{1}{10}$, $c_0 + \frac{2}{10}, \ldots, c_0 + \frac{9}{10}$. The point x must then belong to at least one of the closed subintervals of I_0 (possibly to two adjacent ones if x is one of the points of subdivision). In other words, there is a digit c_1 (that is, one of the integers 0, 1, 2, . . . , 9) such that x belongs to the closed interval I_1 given by

$$c_0 + \tfrac{1}{10}c_1 \leq x \leq c_0 + \tfrac{1}{10}c_1 + \tfrac{1}{10}.$$

Dividing I_1 in turn into ten equal parts, we find a digit c_2 such that x lies in the interval I_2 given by

$$c_0 + \tfrac{1}{10}c_1 + \tfrac{1}{100}c_2 \leq x \leq c_0 + \tfrac{1}{10}c_1 + \tfrac{1}{100}c_2 + \tfrac{1}{100}.$$

We repeat this process. After n steps x is confined to an interval I_n given by

$$c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n \leq x \leq c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n + \frac{1}{10^n},$$

where $c_1, c_2, \ldots$ are all digits. The interval I_n has length $1/10^n$, which tends to zero for increasing n. It is clear that the I_n form a nested set of intervals, and hence that x is determined uniquely by the I_n. Since the I_n are known, once the numbers $c_0, c_1, c_2, \ldots$ are given we find that an arbitrary real number can be described completely by an infinite sequence of integers $c_0, c_1, c_2, \ldots$, where all except the first are digits, having values from zero to nine only. In ordinary decimal notation the connection between x and $c_0, c_1, c_2, \ldots$ is indicated by writing

$$x = c_0 + 0.c_1c_2c_3 \cdots.$$

(Usually, the integer c_0 itself is also written in decimal notation if c_0 is positive.) Conversely, by the axiom of continuity, every such expression denoting an infinite decimal fraction represents a real number.

It is possible that there are two different decimal representations of the same number; for example,

$$1 = 0.99999 \cdots = 1.00000 \cdots.$$

In our construction the integer c_0 is determined uniquely by x unless x itself is an integer. In that case we could choose either $c_0 = x$ or $c_0 = x - 1$. Once a choice has been made the digit c_1 is unique unless x is one of the new points subdividing I_0 into ten equal parts. Continuing we find that c_0 and all c_k are determined uniquely by x unless x occurs as a point of subdivision at some stage. If this should happen for the first time at the nth stage, then

$$x = c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n,$$

where $c_1, c_2, \ldots, c_n$ are digits and where $c_n > 0$, since otherwise x would have been a point of subdivision at an earlier stage. It follows that I_{n+1} is either the interval $[x, x + 1/10^{n+1}]$ or the interval $[x - 1/10^{n+1}, x]$. In the first case x will be the left-hand end point of all later intervals $I_{n+2}, I_{n+3}, \ldots$, and in the second case, the right-hand end point. We are then led either to the decimal representation

$$x = c_0 + 0.c_1c_2 \cdots c_n000 \cdots$$

or the representation

$$x = c_0 + 0.c_1c_2 \cdots (c_n - 1)99999 \cdots.$$

Hence the only case in which an ambiguity can arise is for a rational number x which can be written as a fraction having a power of ten for its denominator. We can eliminate even this ambiguity by excluding decimal representations in which all digits from a certain point on are nines.

In the decimal representation of real numbers the special role played by the number ten is purely incidental. The only evident reason for the widespread use of the decimal system is the ease of counting by tens on our fingers (digits). Any integer p greater than one can serve equally well. We could use p equal subdivisions at each stage. A real number x would then be represented in the form

$$x = c_0 + 0.c_1c_2c_3 \cdots,$$

where c_0 is an integer, and now $c_1, c_2, \ldots$ have one of the values $0, 1, 2, \ldots, p - 1$. This representation again characterizes x by a nested set of intervals, namely

$$c_0 + \frac{1}{p}c_1 + \cdots + \frac{1}{p^n}c_n \leq x \leq c_0 + \frac{1}{p}c_1 + \cdots + \frac{1}{p^n}c_n + \frac{1}{p^n}.$$

If x is positive or zero, the integer c_0 is also positive or zero and c_0 itself has a finite expansion of the form

$$c_0 = d_0 + pd_1 + p^2d_2 + \cdots + p^kd_k,$$

where $d_0, d_1, \ldots, d_k$ take one of the values $0, 1, \ldots, p - 1$. The complete representation of x "to the base p" takes the form

$$x = d_kd_{k-1} \cdots d_1d_0.c_1c_2c_3 \cdots.$$

If x is negative, we may use this kind of representation for $-x$.

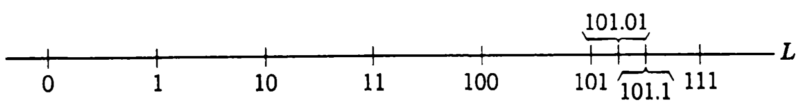

Figure 1.4 The fraction $\frac{21}{4}$ in the binary system.

Bases other than 10 have actually been used extensively. Following the lead of the ancient Babylonians, astronomers for many centuries consistently represented numbers as "sexagesimal" fractions with $p = 60$ as the base.

Binary Representation. The "binary" system with the base $p = 2$ has special theoretical interest and is useful in the logical design of computing machines. In the binary system the digits have only two possible values, zero and one. The number $\frac{21}{4}$, for example, would be written 101.01 corresponding to the formula

$$\frac{21}{4} = 2^2 \cdot 1 + 2^1 \cdot 0 + 1 \cdot 1 + \frac{1}{2} \cdot 0 + \frac{1}{2^2} \cdot 1 \qquad \text{(see Fig. 1.4).}$$

Calculating with Real Numbers. Although the definition of real numbers and their infinite decimal or binary representations, etc., are straightforward, it may not seem obvious that one can operate with the number continuum exactly as with rational numbers, performing the rational operations and retaining the laws of arithmetic, such as the associative, the commutative, and the distributive laws. The proof is simple, although somewhat tedious. Instead of impeding the way to the live substance of analysis by taking up the question here, we shall accept temporarily the possibility of ordinary arithmetic calculation with the real numbers. A deeper understanding of the logical structure underlying the number concept will come when we discover the idea of limit and its implications. (See the Supplement to this chapter, p. 89.)

d. Definition of Neighborhood

Not only the rational operations but also order relations or inequalities for real numbers obey the same rules as for the rational numbers.

Pairs of real numbers a and b with $a < b$ again give rise to closed intervals $[a, b]$ (given by $a \leq x \leq b$) and open intervals (a, b) (given by $a < x < b$). Frequently we shall be led to associate with a point x_0 the various open intervals that contain that point or specifically have it as center, which we shall call *neighborhoods* of the point. More precisely, for any positive ϵ the *ϵ-neighborhood* of the point x_0 consists of the values x for which $x_0 - \epsilon < x < x_0 + \epsilon$, that is, it is the interval $(x_0 - \epsilon, x_0 + \epsilon)$. Any open interval (a, b) containing a point x_0 always also contains a whole neighborhood of x_0.

Having defined intervals with real end points we can now form *nested sequences* of intervals using the same definition as in the case of rational end points. It is most important for the logical consistency of calculus that for any nested sequence of intervals with real end points there is a real number contained in all of them. (See Supplement, p. 95.)

e. Inequalities

Basic Rules

Inequalities play a far larger role in higher mathematics than in elementary mathematics. Often the precise value of a quantity x is difficult to determine, whereas it may be easy to make an estimate of x, that is, to show that x is greater than some known quantity a and less than some other quantity b. For many purposes, only the information contained in such an estimate of x is significant. We shall therefore briefly recall some of the elementary rules about inequalities.

The basic fact is that the sum and product of two positive real numbers are again positive; that is, if $a > 0$ and $b > 0$, then $a + b > 0$