Chapter One

The Tools of Calculus

The complexity of Nature has led to the existence of various sciences that consider the same natural objects using different tools and approaches. An expert in the physics of solids may find it hard to communicate with an expert in the mechanics of solids; even between these closely related subjects we find significant differences in mathematical tools, terms, and viewpoints taken toward the objects of investigation. The physicist and engineer, however, do share a bit of common background: the tools of mathematical physics. These have evolved during the long history of our civilization.

The heart of any physical theory — say, mechanics or electrodynamics — is a collection of main ideas expressed in terms of some particular wording. The next layer consists of mathematical formulation of these main ideas. It is interesting to note that mathematical formulations can be both broader and narrower than word statements. Wording, especially if left somewhat "fuzzy," is often capable of a wider range of application because it skirts particular cases where additional restrictions would be imperative. On the other hand, mathematical studies often yield results and ideas that are important in practice — for example, system traits such as energy and entropy.

Among the most mathematical of the disciplines within physics is mechanics. At first glance other branches such as quantum mechanics or field theory might seem more sophisticated, but the influence of mechanics on the rest of physics has been profound. Its main ideas, although they reflect our everyday experience, are deep and complicated. The models of mechanics and the mathematical tools that have been developed for the solution of mechanical problems find application in many other mathematical sciences. The tools, in particular, are now regarded as an important part of mathematics as well. Indeed, the relationship between mechanics and mathematics has become so tight that it is possible to consider mechanics as a branch of mathematics (although much of mechanics lacks the formalized structure of pure mathematics).

In this book we shall explore the role of mathematics in the development of mechanics. The historical pattern of interaction between these two sciences may yield a glimpse into the future development of certain other fields of knowledge (e.g., biology) in which mathematical rigor currently plays a less fundamental role.

Our use of the term "mechanics" will include both "classical mechanics" and "continuum mechanics." The former treats problems in the statics and dynamics of rigid bodies, while the latter treats the motions, deformations, and stresses of bodies in cases where the details of atomic structure can be neglected. Thus mechanics describes the behavior of a great many familiar things. It can further account for the influence of phenomena previously not considered under the heading of mechanics: a mechanical body can exhibit magnetic properties, for example. Mechanicists developed simple but effective mathematical models of real objects used in engineering practice. These include models of beams, plates, shells, linearly elastic bodies, and ideal liquids. The development of mathematical tools for the solution of specific problems was done in parallel with this, and often by the very same persons. Early on, nobody tried to divide mathematics into pure and applied portions. There was a certain unity between mathematics and physics, and impressive amounts of progress were made on both fronts — even though the number of "research scientists" was considerably smaller than it is today. We know Sir Isaac Newton (1642–1727), for example, as both a great physicist and mathematician. The names Euler, Bernoulli, and Lagrange also summon images of first-rate mathematicians without whom much of continuum mechanics may not have survived. These historical icons, along with many others, developed new mathematical tools not only out of pure mathematical curiosity: they were also experts in applications, and knew what they needed in order to solve important practical problems. They understood the directions that mathematics needed to take during their lifetimes. Augustin-Louis Cauchy (1789–1857), whose name is encountered in any calculus textbook, introduced a key notion in continuum mechanics: that of the stress tensor. He elaborated it by exploiting the same tools and ideas he used to solve problems in pure mathematics. Many important branches of mathematics appeared in response to the very real needs of engineering and the other applied sciences.

Because a consideration of mathematical tools and how they are transformed in mechanical modeling is one of our goals, it makes sense to begin with a discussion of elementary calculus. Indeed, the ideas laid down in mechanical modeling are the same as those laid down in mathematics.

1.1 Is Mathematical Proof Necessary?

Among the most wonderful notions ever elaborated by mankind was that of number. This notion opened the door to comparisons that were previously impossible. Two men and two apples now had something in common: both of these sets could be mentally placed into correspondence with the same set of two fingers on one's hand. Many languages still contain a phrase along the lines of "... as many as the fingers on both of my hands." The sense of number is not exclusively human: some talented crows appear to have it as well. If such a crow is shown two sets of beans, one containing three beans and the other containing four, the bird can choose the larger set without error. It turns out that a crow cannot distinguish between sets of seven and eight beans, but even humans can normally grasp no more than six items at a glance and must count systematically after that. The notion of abstract number was brought into common use not so long ago (historically speaking). The necessity to count abstract quantities gave rise to methods called "algebra." Engagement in trade forced people to compare not only sets of discrete items, but tracts of land, quantities of liquid, and so forth. This led to geometry as the science by which continuous quantities could be compared. People elaborated the notions of area and volume; they had to compare pieces of land of various shapes, and therefore had to consider the main geometrical figures in the plane and in space and find their measures.

The ancient Egyptians and Babylonians could find the area of a rectangle and even a triangle. They had no formulas, so a solution was given as an *algorithm* describing which measurements must be taken and what must be done with the results. The methods of calculation were probably discovered during extensive testing: various areas were covered with seeds, which were then counted and the results compared with the output of some algorithm. Of course, this is only an assumption because all we have are the algorithms themselves, found on Babylonian clay tablets and similar sources; at that time nobody was interested in describing how they made such discoveries. Modern applied mathematical sciences also make extensive use of algorithms. Again, an algorithm is merely a precise description of the actions necessary to find some desired value. In this way, modern mathematics is similar to that of the Babylonians. It is possible that ancient mathematicians were even more careful than modern applied mathematicians: ancient algorithms were tested thoroughly, whereas today many algorithms are based largely on intuition (which is not so bad if the intuition is good).

The problem of comparison was the beginning of all mathematics. Of course, mathematics has evolved into a great many distant branches. But the common approach was developed by the ancient Greeks. The Greeks decided that it was necessary to lend support to the algorithms for the calculation of lengths, areas, and volumes known from even more ancient times. The word of an authority figure was no longer deemed sufficient: the Greeks also decided to require *proof*.

Interactions between civilized people are based on rules that all or most people consider to be valid. It is impossible to force people to do anything all the time; normal social relations are based on the act of convincing people to behave in certain ways. This in turn requires a convincing argument. This viewpoint culminated in ancient Greece where the laws of democracy forced people to learn the art of convincing others. The art became so important that certain persons began a systematic study of its main elements so that they could teach these to other people. Discussion itself was analyzed with the goal of figuring out how to win. Out of this came the main elements of logic and the standard modes of reasoning. Eventually some teachers left the practical realm of application of this knowledge and began to study Nature using the same approach. This led them to quite abstract notions. They developed an understanding of various effects, and tried to understand why things happen. The thinkers who had the most advanced tools, and who applied them in all situations, became known as philosophers. Among the most important realms was that which we now call *geometry*.

Ancient geometry combined the whole of the mathematics of its time. It was regarded as the ultimate abstraction and capable of explaining everything in Nature. Moreover, some people like Pythagoras expected that full knowledge of mathematical laws would give one power over the whole world. Numbers were thought to govern everything — learn their laws and you become king of the universe! This became a slogan for a certain direction of Greek philosophy that evolved into a kind of religion. The art of convincing argument was applied to the most abstract subject of the time: mathematics. Mathematics became full of abstract reasoning and was considered by the ancient Greeks to stand at the pinnacle of philosophy. This point of view came down to premodern times. In certain Russian universities prior to the October Revolution, for example, mathematics was relegated to departments of philosophy. This is why the Ph.D., the Doctor of Philosophy, is the title given to mathematicians and physicists.

To understand what the Greeks brought into mathematics, it is helpful to take a look at how their discussions were arranged. The dialogues of Socrates, as described by Plato, consisted of long chains of deductive reasoning. Each statement followed from the one before, and any chain of statements began with one that his adversary accepted as indisputable. We oversimplify Socrates' method a bit here, but the scheme we describe is the one that was introduced into mathematics. Mathematics became more than a set of practical algorithms; it was supplied with the tools one needed to argue that the algorithms were correct. The peak of this approach was presented in Euclid's *Elements*, which until recently was a standard text in geometry. Euclid (c. 300 B.C.) collected much of the geometry known in his time. His approach was copied by many other branches of mathematics and, when logical, by various other sciences.

What is the structure of the *Elements*? Euclid regarded certain central notions of geometry as self-evident. He did not try to define a point or a straight line, but only described their properties. These properties were taken as universally accepted. For example, a straight line gave the shortest distance between two points. The most evident of the basic assertions were called *axioms*, and the rest were called *postulates*.

With the axioms and postulates for the main elements (points, straight lines, and planes) in hand, the main figures (triangles, polygons, circles) can be introduced, and a study of the relations between these can begin. Not all "self-evident" statements remained so. Euclid's famous fifth postulate states that through a point outside a straight line on the plane it is possible to draw a parallel straight line, and that this line is unique. Later geometricians tried to prove this as a theorem. In the nineteenth century it was found to be an independent statement that might be true or untrue in real space — we still do not fully understand the real geometry of our space. So this postulate turned out to be nontrivial, and modern geometry regards it as an axiom for the so-called Euclidean geometry. But there are geometries with other sets of axioms. In Lobachevsky's geometry, through a point on a plane there are infinitely many straight lines parallel to a given one; in Riemannian geometry, it is taken for granted that on the plane there are *no* lines parallel to a given one. This results in some differences in theorems between the various branches of geometry. In non-Euclidean geometries, the interior angles of a triangle do not sum to 180°. A consequence is that we cannot precisely determine what qualifies as a straight line or a plane. We have only an image of a straight line as the path along which a ray of light reaches us from a distant star. But such a path is not straight: its course is altered by gravitation. So we can point to nothing in our environment that is a straight line. The same holds for the idea of a plane, notwithstanding the fact that children are told that the surface of the table in front of them is a portion of a plane.

The relations between geometrical figures were formulated as theorems, lemmas, corollaries, and so on, among which it is hard to establish strict distinctions. But all the central results were called *theorems*. These are formulated as certain conditions on the geometric figures, and the resulting consequences. All the theorems were *proved*, which meant that all the consequences were justified using the axioms and theorems proved earlier. This was a great achievement for a comparatively small group of people. In the *Elements*, even quadratic equations were solved through the use of geometrical transformations. This book was the standard of mathematicians for many years, and modern students who dislike similar logical constructions may consider themselves victims of the ancient Greeks.

Today, mathematics remains at the center of science. But it has developed along with everything else. Arguments considered rigorous by the ancient Greeks are sometimes considered to be flimsy nowadays. In 1899, an attempt was made to construct an absolutely rigorous geometry. In his famous *Foundations of Geometry*, David Hilbert (1862–1943) took the Greek formalization to an extreme level. Hilbert stated that there are to be undefined main elements — points, straight lines, and planes — and between these, some relations should be determined axiomatically. Then all the statements of geometry should be derived from these axioms independently of what one might regard as a point, line, or plane. The reader can view these terms in any fashion he or she wishes: students as points, desks as lines, and cars as planes. But despite a complete misunderstanding of these main elements, he or she should arrive at the same relations between them as would someone who had a better viewpoint (we say "better" because the notions are undefined, so nobody can have a perfect understanding). The geometry of Hilbert contains many axioms and is much more formal than that of Euclid. Later, failed attempts were made to prove consistency among the full system of geometrical axioms. The first attempt to axiomatize geometry spurred many mathematicians to try to axiomatize all of science. Mathematics was the pioneer of all rigorous study, and it has been said that the only real sciences are those that employ mathematics. (Should we believe such a statement?) Hilbert's book inspired other mathematicians to develop all of mathematics as an axiomatic–deductive science, where all facts are deduced from a few basic statements that have been formulated explicitly. Attempts have been made to introduce the axiomatic approach into mechanics and physics, but these have not been very successful. The totally axiomatic approach in mathematics was buried by Kurt Gödel (1906–1978), who proved two famous (and, in a sense, disappointing) theorems:

1. The *first inconsistency theorem* states that any science based on the axiomatic approach contains undecidable statements — theorems that could not be proved or disproved. Therefore such statements can be taken as independent axioms of any version of the theory: positive or negative. This means that any axiom system is incomplete if we wish to find all the relations between some elements for which the axioms were formulated.
2. The *second inconsistency theorem* states that it is impossible to show consistency of a set of axioms using the formalization of the system itself; only by using a stronger system of axioms can we establish such consistency. This means that even the simplest axiom system describing arithmetic cannot be checked for consistency using only the tools of arithmetic.

This did not mean the end of mathematics. But it did bring some indefiniteness into the strictest of the sciences. The way in which the ancient Greeks proved mathematical statements remains the most reliable method available, and students will continue to study proofs of the Pythagorean theorem for years to come.

1.2 Abstraction, Understanding, Infinity

A small boy who announces that he fails to understand the concept of "one" is not necessarily being silly. There is no such thing as "one" in Nature. We can point to one finger or one pie, of course, but not to "one" as an independent entity. The concept of "one" is quite fluid: we can use it to refer to a whole pie or to a single piece of that pie. So the boy may have good reason to be perplexed. It is probably fair to say that we become accustomed to abstract notions more than we actually understand them. Two high school trigonometry students may exhibit identical skill levels and yet disagree profoundly as to whether the subject is understandable. This sort of difference is essentially psychological, and in many cases relates directly to how long we have received exposure to an abstract concept. Young persons seem to be especially well equipped to accept new notions and to use them with ease. As we get older we seem to require much more time to decide that something is clear, understandable, and easy.

One abstract notion that takes a while to accept is that of infinity. We overuse the word "infinite," even applying it when we are angry about having to stand in line too long. But infinity is a troublesome concept: we are at a loss to really visualize it, yet seem to be able to vaguely intuit more than one version of it. One could write volumes about how to imagine infinity. However, the thicker the book, the less clear everything would become; our understanding should spring from the simplest possible ideas, and proceed systematically to the more complex.

Centuries ago, there were attempts to treat infinite objects in the same way as finite geometrical objects. In this way, the properties of finite objects were attributed to infinite objects. An "actual infinity" was introduced and treated as though it were an ordinary number. Many interesting arguments were constructed. For example, one could "prove" that any sector of a plane (i.e., the portion lying between two rays emanating from a common point) is larger than any strip (i.e., a portion lying between two parallel lines). The argument was simple: an infinite number of strips will be needed to cover the plane, whereas a finite number of sectors will suffice. So the area of the sector must be larger. This and many similar "proofs" introduced paradoxes into mathematics and forced people to reconsider the concept of infinity. What finally remained was a "potential infinity" that could be only approached using processes involving finite shapes or numbers.

Let us see how infinity enters elementary mathematics. We should take the simplest object we can and try to understand how infinity comes into it. This object is the set of positive integers. It contains no largest element: any positive integer can be increased by one to yield a new positive integer, so we say that the positive integers stretch on to infinity. This is an infinite set with the simplest structure. We can compare the sizes of two sets by using the idea of a one-to-one correspondence. We say that two sets have the *same cardinality* if they can be placed in such a correspondence, regardless of the nature of the elements themselves. The positive integers and its various subsets are useful for such comparisons. This is essentially what we do when we count the elements of a set: we say that a set has n elements if we can pair each element uniquely with an integer from the set $\{1, 2, 3, \ldots, n\}$. When no value of n suffices for this, we say that our set is infinite. The set of positive integers has the least possible cardinality of all infinite sets, and any set with the same cardinality is said to be *countable*. Thus, each member of a countable set can be labeled with a unique positive integer. It is clear that if we put two countable sets A and B together (i.e., form the *union* of the sets), we obtain a new countable set. The elements of the new set can be counted as follows: the first member of A is paired with the integer 1, the first member of B is paired with the integer 2, the second member of A is paired with the integer 3, the second member of B is paired with the integer 4, and so on. Thus, from the viewpoint of cardinality, the union of A and B is no larger than either of these individual sets. The same thing happens when we consider the union of any finite number of countable sets: the result is countable. Moreover, the union of countably many sets is always countable. (The reader could try to demonstrate this by showing how the elements of the union can be renumbered so that each is paired uniquely with an integer.) From this, it follows that the set of all rational numbers is countable. From one perspective, the rational points cover the number line densely, so that in any neighborhood of any point we find infinitely many rational points; however, in the sense stated above, the rationals are no more numerous than the positive integers — which are quite rare on the same axis. Of course, this is a bit bothersome. Another interesting question is whether there exists a set with cardinality higher than the positive integers. Mathematicians have shown that the set of points contained in the interval $[0, 1]$ is of this type: these points are too numerous to be paired with the positive integers. We say that $[0, 1]$ has the cardinality of the *continuum*. It turns out that the same cardinality is shared by the entire number line, the plane, and three-dimensional space as well. Thus we have come to consider another kind of infinity. We shall not touch on infinities of even higher order, though they exist; the curious reader can consult a textbook on set theory for more information. We should, however, ask whether sets exist with cardinalities intermediate between those of the positive integers and the continuum. The assumption that they do, it turns out, is a fully independent axiom of arithmetic. The stipulation of this axiom (or lack thereof) will not affect the results on which we so often depend. This is a confirmation of Gödel's famous theorem, because it demonstrates the incompleteness of the axioms of arithmetic — the simplest (and seemingly nicest) axiom system in mathematics.

Our easy familiarity with the positive integers leads us to think that we should be able to perform countably many actions ourselves. Indeed, we effectively do this when we sum an infinite series via a limit passage. After becoming bored with such things, mathematicians decided to attempt actions as numerous as the points of the continuum. But this led to serious paradoxes. It was found, for example, that one could carve up a sphere of unit radius into a great many pieces, alter their shapes without changing their volumes, and then collect the new pieces into a sphere of radius 1,000,000 with no holes in the latter (the Banach–Tarski Paradox[1])! Hence the problem of infinity was considered more carefully. This problem is the central point of calculus and other parts of analysis that relate to differentiation and integration, but there, limit passages are presented in such a way that only finite quantities are involved at each step. This explains why modern students can quickly feel secure in their necessary dealings with the subject.

1.3 Irrational Numbers

When we purchase 1 meter of cloth, we may actually receive 1.005 meters. Of course, we probably will not care; for our purposes the tiny bit of extra cloth will not matter. We might even refer to two pieces of cloth having

[1]See *The Banach–Tarski Paradox* by Stan Wagon (Cambridge: Cambridge University Press, 1993).

equal widths, with lengths of 1 m and 1.005 m, respectively, as being of "equal" size. This kind of fuzzy thinking has its place in everyday life. But it was also characteristic of very ancient mathematics.

The Greeks refused to accept this approach. In their view equality was *absolute*. Two triangles were considered equal only when the first could be exactly superimposed upon the second; any difference, no matter how small, made them unequal. Now, it may be hard to imagine a mathematics based on "approximate" equalities, but things could have developed that way. We would have lost algebra and the strict results of geometry, but might have gained something else. Let us avoid speculation, however, and consider what actually did happen instead.

In geometry, one of the simplest and most crucial problems was the measurement of the length of a straight segment. The length of a given segment should be compared with the length of a standard segment that can be said to possess unit length. Any physics student is aware that the meter length was determined by the length of a standard metal rod kept under lock and key somewhere by a standards organization. We measure everything in terms of the length of this rod. But although the rod is maintained under controlled environmental conditions, its length will vary a bit due to tiny changes in air temperature, pressure, and so on, in its holding chamber. These changes may be small but they are almost certainly not zero, and are definitely an issue when we are trying to determine whether two lengths are absolutely equal. Even subtle changes in electromagnetic fields caused by solar activity will alter the length of the standard meter bar.

Retaining the viewpoint of the ancient Greeks who introduced absolute equality, we may or may not even be able to find two moments in time when the lengths of this same bar will be equal.

Thus, the proudly absolute Greeks proposed a way in which the length of an arbitrary segment could be measured in terms of a unit-length standard segment. Our modern adaptation is as follows. We place the standard segment next to the unknown segment with two of the ends flush. If the unknown segment is longer, then we mark the position of the other endpoint of the standard segment on the unknown segment, and repeat this procedure until the portion of the unknown segment that remains is less than the length of the standard segment. In this way, we find the integer part of the unknown length. We next divide the standard segment into several equal parts, say ten, and repeat the procedure using one-tenth of the standard segment. We can continue to subdivide the standard, and in this way determine the unknown length to any desired degree of accuracy.

What the Greek mathematicians did was actually a bit different. They first supposed it possible to divide a standard segment into several equal parts, say n parts, in such a way that the unknown segment is composed of an integer number of such parts. If this integer number is m, then in modern notation the unknown length is m/n. At first, the Greeks were sure this would always work. Their certainty was rooted in their belief that the structure of the world was ideal (and many modern scientists share this view). It is interesting to note that the Greek philosophers held notions akin to our present knowledge of the atomic structure of materials. They did, however, have a different idea of what an atom would look like if one could be seen. Atoms were thought to be objects that completely filled in the exterior framework of a material body. Because all atoms of a certain material had to be identical, and these had to fill in the body completely without gaps, atoms were supposed to take some ideal form — possibly that of a perfect polyhedron. This formed the basis of a mathematical/religious belief that integer numbers govern the world. And so the ancient mathematicians thought they could measure the proportions of any perfect figure through the use of integers only.

This belief persisted until it failed for the square. If a square has sides of length 1, a quick calculation using the Pythagorean theorem shows that the diagonal has a length, in modern terms, of $\sqrt{2}$. If the above method of measuring lengths were valid, this length would have to equal m/n for some integers m, n. We will demonstrate that this cannot be the case. Let us first assume that m and n have no common integer factors: that is, that all factors common to the numerator and denominator of the fraction m/n have been cancelled already. Let us now suppose that there are integers m, n such that

$$\sqrt{2} = m/n.$$

Squaring both sides of the equality, we get

$$2 = m^2/n^2,$$

or

$$m^2 = 2n^2.$$

Now, m^2 cannot have 2 as a divisor unless m does also. So there must be another integer p such that $m = 2p$. Substituting this into the last equality, we have

$$4p^2 = 2n^2,$$

or

$$n^2 = 2p^2.$$

But this means that n also has 2 as a divisor, and can therefore be expressed as $n = 2q$ for some integer q. This, of course, contradicts our assumption that m and n have no common integer factors. Hence, $\sqrt{2}$ cannot be presented in the form of a ratio m/n in which m, n are integers; we now say that $\sqrt{2}$ is not a *rational number*. This spelled the end for some of the old Greek ideals, but brought into play a new class of numbers called *irrational numbers*.

What is an irrational number? The symbol $\sqrt{2}$ is nothing more than a label unless we can say precisely what this number is. And this we actually cannot do, but we can calculate the value of $\sqrt{2}$ to within any desired accuracy. We can find as many decimal digits of this number as we wish, but can never state its exact value except by giving the label $\sqrt{2}$. Irrational numbers are not exceptional cases. In higher mathematics, it is shown that the situation is just the opposite: rational numbers are the exceptions, and in a certain technical sense the number of rationals is negligible in comparison with the number of irrationals. So we should elaborate on how to describe an irrational number precisely. There are two ways of doing this. The first is simply to label the number. This is possible for some important irrational numbers such as π and e, but not for all of them. The second is to describe how to pin down an irrational number more and more precisely. Thus we can say that an irrational number is the result of some infinite process of approximation. In this way we arrive at the notion of *limit*. The Greeks used this notion in an intuitive manner, but the strict version had to wait until much later — when it could form the backdrop for the calculus.

1.4 What Is a Limit?

Let us pursue the question of how irrational numbers are defined. This leads us to the more general idea of limit. We will need the notion of distance from geometry. We introduce a horizontal line (*axis*) on which we can mark a zero point, a unit distance away from zero, and a positive direction (say, to the right). Using these, we can mark a point corresponding to a rational number m/n. This is the point whose distance from zero is $|m/n|$; it falls to the right of zero if m/n is positive or to the left of zero if m/n is negative.

In this way, we can mark enormously many points on the numerical axis, but some points (e.g., $\sqrt{2}$) must remain unmarked. Note that we have in mind a nice picture of a straight line. But remember that we really do not know what a straight line is: we only know some of its properties. In geometry it is often said that a straight line is a portion of a circle having infinite radius. This notion can be bothersome if we try to envision the full circle; it is reasonable, however, if we concentrate only on a certain portion of interest, and the idea does work nicely in doing calculations! Another question worth considering is whether we could ever draw even a segment of a truly straight line, or come up with some other suitable material representation. If we keep in mind that all materials are made up of separate atoms, then it becomes hard to imagine constructing a continuous straight segment. If we add that some modern physical theories regard space as having a cell structure, then drawing a straight line becomes more than problematic. However, idealization is still a convenient way to investigate relations between real-world objects. To precisely define the concept of an irrational number, we can apply the idea of successive approximations as introduced in the previous section. We will modify that idea as follows.

To locate any real number a, whether irrational or rational, we begin by bracketing it between two points on either side. Let us write

$$a_0^- \le a \le a_0^+,$$

where the *lower* and *upper bounds* $a_0^\pm$ are rational. We can refine our approximation of a by giving another rational pair $a_1^\pm$, between which a is squeezed more tightly: that is, we want

$$a_1^- \le a \le a_1^+,$$

where $a_0^- \le a_1^-$ and $a_1^+ \le a_0^+$. Repeating this process, we can bracket a more and more precisely, and thereby think of approximating it more and more closely. Our construction is such that the kth approximating segment $[a_k^-, a_k^+]$ lies within all the previous segments $[a_i^-, a_i^+]$, $i = 0, 1, \ldots, k-1$, and the lengths $|a_k^+ - a_k^-|$ decrease toward zero as k increases indefinitely.

It seems evident that some point on the axis will belong to each of the approximating segments $[a_k^-, a_k^+]$, and that this point will be unique. Unfortunately, because we know little more than the ancient Greeks knew about straight lines, we do not know anything about the "local structure" of the line and must therefore take this "evident" statement for granted. (The reader might consider proposing some alternatives: perhaps there are no such points, or two, or infinitely many?) This is an example of an axiom: a statement that cannot be proved but which can be accepted so that argumentation can proceed. So we hereby introduce an axiom of nested segments: there is a unique point that belongs to each of a sequence of nested segments whose lengths tend ever toward zero.

Despite the fact that things are looking evident and even fairly scientific now, we have managed to slip in another fuzzy notion. What does the phrase "tend toward zero" mean? A precise definition was proposed by Cauchy, the French mathematician so prolific that the flow of his papers overwhelmed the mathematical journals of his time (he eventually had to publish his own journal). When we hear such impressive sounding names we are prone to imagine persons whose lives were led in a sophisticated and academic manner, and who from time to time published important results that might immediately come to be called Cauchy's theorem or Lagrange's theorem. But in ordinary life, great men sometimes behave as badly as anyone else. Cauchy's results were so numerous that, as the story goes, another of his great contemporaries characterized Cauchy as having "mathematical diarrhea." In return, Cauchy claimed that his contemporary had a case of "mathematical constipation." We shall meet Cauchy again on future pages; he was truly great and originated many important ideas. One of these was quite nontrivial: how to describe infinite processes in terms of finite arguments — again, the notion of limit. Despite our desire to be extremely advanced in comparison with other animals, we find ourselves quite restricted when we attempt to do anything exactly. We cannot tell whether a bag contains eight or nine apples without counting them, just as a crow could not (there are exceptional persons such as that portrayed by the main character of the movie *Rain Man*, but these persons merely possess higher ceilings for making immediate counts). We definitely cannot perform infinitely many operations even if we declare that we can. So we need a way to confirm that all has gone well in a process having an infinite number of steps, even if we can carry out only finitely many. This is the central point of approximation theory, and thus the theory of limits.

We consider the simple notion of sequence limit. The notion of sequence is well understood. For example, a sequence with terms $x_n = 1/n$ looks like

$$1\,, \quad 1/2\,, \quad 1/3\,, \quad \ldots, \quad 1/n\,, \quad \ldots.$$

If we consider the term numbered $n = 100$, we get $x_{100} = 1/100$. For $n = 1000$ we get $1/1000$, which is closer to zero; for $n = 1000000$ we get $1/1000000$, which is even closer. The higher the index n, the closer the sequence term is to zero. Thus zero plays a special role for this sequence and is called the *limit*. In our example, the sequence terms decrease monotonically, but this does not always happen, and we can phrase the limit idea more generally by saying something like "the greater the index, the closer the terms get to the limit." Cauchy proposed the following definition.

Definition 1.1 The number a is the *limit* of the sequence $\{x_n\}$ as $n \to \infty$ if for any $\varepsilon > 0$ there is a number N, dependent on ε, such that whenever $n > N$, we get $|x_n - a| < \varepsilon$. A sequence having a limit is called *convergent*, and one having no limit is called *nonconvergent*. If a is the limit of $\{x_n\}$, then we write

$$a = \lim_{n\to\infty} x_n,$$

or $x_n \to a$ as $n \to \infty$.

Note that if we consider a stationary sequence — that is, a sequence whose elements $x_n = a$ for all n — then by definition a is the limit. This supports our view of the limit as a value approached more and more closely as the index n increases.

Let us carefully consider the limit definition from several points of view, because it is the central definition of the calculus and of continuum physics. First, there is the requirement "for any $\varepsilon > 0$". The word "any" is dangerous in mathematics. In this context, it means that if we are to verify that a sequence has limit a, we must try *all* positive values of ε without exception; for each $\varepsilon > 0$, we must be able to find the number N depending on ε mentioned next.

At first glance, we appear to have gained nothing: we started with an infinite process of either convergence or nonconvergence to a limit, and have replaced it by another, apparently infinite, process that forces us to seek a suitable N for every possible positive number ε. Note that it can be described as follows. We take some sequence of values ε_k tending toward zero, and for each of these find an $N(\varepsilon_k)$. It is clear that if we find $N(\varepsilon_1)$ then this N can be taken for all $\varepsilon > \varepsilon_1$. Thus we need not consider the continuous set of all $\varepsilon > 0$. Moreover, although the definition states that we should verify the property for all ε, our serious work begins with those that are close to zero. But the new process need not be infinite: we can find an analytical dependence of N on ε by solving the inequality

$$|x_n - a| < \varepsilon$$

with respect to the integer n, and then see whether the solution region contains some interval $[N, \infty)$. If so, then a is the limit of the sequence. If there is an $\varepsilon > 0$ for which there is no such N, then the sequence has no limit. Thus our task does not involve infinitely many checks, but rather the solution of an inequality. This is how the result of an infinite process can be verified through a finite number of actions.

Note that the definition of sequence limit does not give us a way to find the limit a. We must produce a candidate ourselves before we can check it using the definition.

Let us reformulate the definition in a more geometrical way. We rewrite the above inequality in the form

$$|x - a| < \varepsilon.$$

The set of all x that satisfy this is $(a - \varepsilon, a + \varepsilon)$, which is called the ε-*neighborhood* of a on the numerical axis. So, geometrically, when we find the solution $[N, \infty)$ of the above inequality, we find the number N beginning with which all the terms lie in the ε-neighborhood of a. That is, if we remove all terms of the sequence whose indexes are less than or equal to N, then the whole remaining "tail" of the sequence must fall within the ε-neighborhood of a. The number a is the limit if we can do this for $\varepsilon > 0$ without exception. The "tail" idea can be made clearer if we think of the sequence $\{x_n\}$ as a function given for an integer argument n. Figure 1.1 serves to illustrate this viewpoint. The picture extends infinitely far to the right ($n \to \infty$), and we see that convergence or nonconvergence of the sequence depends on what happens to the far right. If we were to drop some of the first terms (even the first few billions of terms, which would still be just the "first few" in comparison with the infinitely many terms present), there would be no effect on convergence or nonconvergence.

We will not consider all the properties of sequence limits, but would like to add one important fact. When a sequence is convergent, then it has one *and only one* limit. This is referred to as *uniqueness* of the limit, and is a property that can be established as a theorem based on the limit definition.

Now our formulation of the axiom for nested segments is exact, and we can return to the problem of defining an irrational number.

We represented each irrational number through some limiting process using nested segments (containing the number) whose lengths tend to zero. In

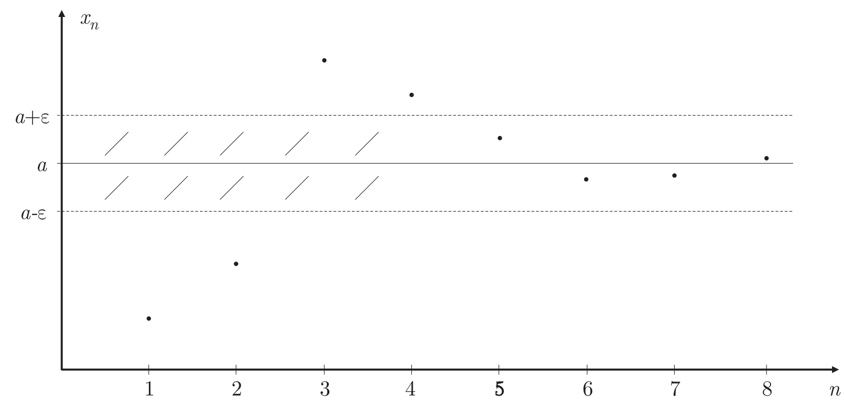

Figure 1.1 One view of sequence convergence. All terms of the sequence after x_4 are inside the band of thickness 2ε (between $a - \varepsilon$ and $a + \varepsilon$), the ε-neighborhood of a. In order to verify that a is actually the limit of $\{x_n\}$, we must verify that a similar picture holds for any positive ε.

fact we can do the same thing for any rational number x, by considering zero-length nested segments $[x, x]$. So each real number, whether rational or irrational, can be defined in a similar mode. Of course, this sort of academic "stuff" would be useless if the arithmetic of such numbers were to give results that disagreed with those of ordinary arithmetic. It turns out that we can employ the usual four arithmetic operations in the same way as we do for fractions. When we add two numbers, whether rational or irrational, the nested approximating segments for these define the ends of segments containing the sum. These segments turn out to be nested as well, and their lengths tend to zero along with the lengths of the original segments that bracketed the operands. In this way we get limit points that define the sum. This can clearly be done for subtraction, multiplication, and division as well. The details can be found in more advanced books. We should also note that our description of an irrational number is not unique: other valid interpretations have been invented, and all have been shown to be equivalent. But the important fact is that we can treat rational and irrational numbers similarly when we do arithmetic and other operations.

1.5 Series

The notion of limit is central in continuum physics. It brings us various mathematical models, at the core of which lie differential or integral equations. These relations are usually derived from geometrical considerations involving infinitesimally small portions of objects. Together with some additional conditions given on the boundary of the object, they form so-called boundary value problems. Because the solutions are frequently given in terms of *series*, we would like to examine this important concept. A series is an "infinite sum," commonly denoted as

$$\sum_{n=1}^{\infty} a_n,$$

which simply means that we must perform the addition

$$a_1 + a_2 + a_3 + a_4 + \cdots$$

step by step to infinity. It is clear that this involves a limit passage. Elementary math students are frightened of series, sometimes refusing to believe that any infinite sum of positive numbers could turn out to be finite. The ancient mathematician-philosophers expressed similar reservations: the famous Zeno (c. 495–430 B.C.), while considering the nature of space and motion, constructed a well-known paradox about the swift runner Achilles, who could never catch a turtle.

The proof was as follows. Between Achilles and the turtle, there is some distance. By the time Achilles covers this distance, the turtle has moved ahead by some distance. By the time Achilles covers this new distance, the turtle has moved ahead by yet another distance. The resulting process of trying to catch the turtle never ends. For Zeno, it was evident that a sum of finite positive quantities must be infinite, which meant that the time necessary for Achilles to catch the turtle must be infinite. Here, Zeno was led to consider a sum of time periods of the form

$$t_1 + t_2 + t_3 + t_4 + \cdots ,$$

which, again, he regarded as obviously infinite.

We shall see that certain sums of this type are infinite, but others can be interpreted in a limit sense. The latter are useful in practice. A common but sometimes overlooked example is the ordinary decimal representation of the number π:

$$\pi = 3.1415926535897 \ldots .$$

This can be viewed as the infinite sum

$$3 + 0.1 + 0.04 + 0.001 + 0.0005 + 0.00009 + 0.000002 + \cdots .$$

We have all used some approximation of π, truncating the series somewhere (usually at 3.14) with full awareness that we would have to include "all of the terms" to get the "exact result." So in some sense it must be possible to "sum" an infinite number of terms and obtain a finite answer. (The Greeks did not have a positional notation for decimal numbers, hence this argument would probably not have convinced them. This way of representing numbers was invented later by the Arabs.) Let us consider what it means to sum a series. Consider a finite sum

$$S_n = a_1 + a_2 + \cdots + a_n.$$

This can approximate the full sum of a series, just as 3.14 can approximate π. Increasing the number of terms by one, we get

$$S_{n+1} = a_1 + a_2 + \cdots + a_n + a_{n+1}.$$

In this way, we get a sequence

$$\{S_n\} = (S_1, S_2, S_3, \ldots, S_n, \ldots)$$

that is no different from the sequences we considered in the previous section. Hence we may consider the problem of existence of the limit for such a sequence. If this limit

$$S = \lim_{n\to\infty} S_n$$

exists, then it is called the sum of the series

$$\sum_{n=1}^{\infty} a_n,$$

and the same symbol $\sum_{n=1}^{\infty} a_n$ is used to denote the sum S as well. In this case, the series is called *convergent*. If the sequence S_n has no limit, then the series is called *nonconvergent*, but the symbol for an infinite sum is still written down formally. In higher mathematics, it is found that there are nonconvergent series whose finite sums S_n yield good approximations to certain functions for which formal infinite sums have been derived. Such series are called *asymptotic expansions*. A well-known example is the Taylor series for a nonanalytic function.

Having seen that irrational numbers can be represented as convergent series, we understand that the series idea is useful and convenient. But this is only the beginning. Many functions can be also represented as convergent series. An example is the expansion into a power series of $\sin x$:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

With this, we can calculate the value of $\sin x$ for small x to good accuracy by using a finite number of terms. Moreover, for $|x| < 1$, the error is less than the last term at which we break off the sum. If we calculate the sum of the first three terms for $x = 0.1$, which is $0.1 - 0.001/6 + 0.00001/120$, then we get $\sin 0.1$ with ten digits of precision.

In past centuries, this and similar series were used to produce tables of functions like $\sin x$, $\cos x$, and so on. The same process is implemented by modern electronic calculators. The speed of computers has made it easier to include more and more terms in an approximation, somewhat alleviating the need to justify the accuracy of each calculation on a theoretical basis. However, the fundamental problem of justifying series convergence remains. We must understand that no result for a finite number of terms can justify the convergence of a series — the situation is the same as with the limit of a sequence.

A few series can be calculated exactly. An example is the series of terms of a decreasing geometric progression:

$$b, bq, bq^2, bq^3, \ldots, bq^n, \ldots \qquad (|q| < 1).$$

The first n terms sum according to the formula

$$S_n = b\frac{1 - q^{n+1}}{1 - q}.$$

It is clear that $\lim_{n\to\infty} q^n = 0$ (the reader can either prove this or become convinced by evaluating a few terms on a calculator), and thus,

$$b + bq + bq^2 + \cdots + bq^n + \cdots = \lim_{n\to\infty} S_n = \lim_{n\to\infty} b\frac{1 - q^{n+1}}{1 - q} = \frac{b}{1 - q}.$$

This series plays a key role in the theory of series convergence, because it leads to the following comparison test.

Theorem 1.2 If a series $\sum_{n=1}^{\infty} a_n$ with positive terms a_n is convergent, then a series $\sum_{n=1}^{\infty} b_n$ with terms such that $|b_n| \le a_n$ is also convergent. If $a_n \le b_n$, and the series $\sum_{n=1}^{\infty} a_n$ is nonconvergent, then the series $\sum_{n=1}^{\infty} b_n$ is also nonconvergent.

The geometric progression is widely used to determine whether a series is convergent or nonconvergent. In the case of convergence, it also affords an estimate of the error due to finite approximation of the series, and can give an indication of the convergence rate. Some of the most capable mathematicians in history have devised series convergence tests. According to Cauchy, we can investigate the behavior of $\sum_{n=1}^{\infty} a_n$ by examining the behavior of $\sqrt[n]{|a_n|}$ as $n \to \infty$. If the limit is less than 1, then the series is convergent; if it is greater than 1, the series is nonconvergent; if it is equal to 1, no conclusion can be drawn. A similar test due to Dirichlet[2] has us take the limit of the ratio a_{n+1}/a_n. The conclusions regarding convergence are the same as in the previous test.

[2]Peter Gustav Lejeune Dirichlet (1805–1859).

The reader can verify that these tests do not yield a definitive result for the series

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots, \qquad 1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots,$$

which are nonconvergent and convergent, respectively. But convergence of the latter is slow and we must retain many terms to get good accuracy.

We will meet further series in the following sections.

1.6 Function Continuity

A sequence can be regarded as a function whose argument runs through the positive integers and whose values are the sequence terms. That is, we may regard $\{x_n\}$ as the set of values of a function $f(n)$ by taking $f(n) = x_n$. But we know that many other functions exist whose arguments can vary over any set on the real axis. We shall now extend the limit concept to functions in general. It is here that the important notion of continuity arises.

Let us briefly review the idea of a real valued function. It is worth doing this because, in continuum mechanics, we will need to consider its generalizations to functions taking vectorial and tensorial values, and even to functions having other functions as arguments; the latter are called *functionals* if their values are real or complex numbers, and *mappings* (or *operators*) if their values are functions or other mathematical entities. We could say the following:

> A function f is a correspondence between two numerical sets X and Y, under which each element of X is paired with at most one element of Y. The set of all $x \in X$ that are paired with some $y \in Y$ is called the *domain* of f and is denoted $D(f)$. The set of all $y \in Y$ that correspond to some $x \in D(f)$ is called the *range* of f and is denoted $R(f)$.

In high school mathematics, we learn about functions for which X is a subset of the real axis, but more general sets of arguments x are common in higher mathematics. The notion of function reflects our broad experience with such relations between many characteristics of objects in the real world.

As usual, we shall not cover all the material given in standard textbooks on calculus, but we will consider the main properties of the tools of calculus — derivatives and integrals — that will be needed in what follows.

We are interested in what happens to the values of a function $y = f(x)$ at a point $x = x_0$. Our experience with graphing simple functions in the plane suggests that most graphs are constructed using continuous curves. These curves may be put together in such a way that the resulting graph has some points at which unpleasant things happen: the graph can approach infinity or possess a sharp break. It is natural to refer to the portions of the graph away from such points as being "continuous." However, this cannot serve as a definition of continuity because some functions are so complex that we cannot draw their graphs. We would first like to understand what happens to the values of a function $f(x)$ as x gets closer and closer to some point x_0. An important case is that in which $f(x)$ is not determined at x_0. Let us bring in the notion of sequence limit as discussed in the previous section. Thus, we consider a sequence $\{x_n\}$ from the domain of $f(x)$, which tends to x_0 but does not take the value x_0. The corresponding sequence of values $\{f(x_n)\}$ may or may not have a limit as $n \to \infty$. If not, we say that $f(x)$ is discontinuous at x_0. Suppose $\{f(x_n)\}$ has a limit y_0. Is this enough to decide that everything is all right at x_0? The answer is no; it is possible that we have chosen a very special sequence. A good example of this is afforded by the function $\sin(1/x)$ considered near $x_0 = 0$. Taking $x_n = 1/(\pi n)$, we get $f(x_n) = 0$ for each n, hence the limit of the sequence of function values is zero; but taking $x_n = 1/(2\pi n + \pi/2)$, we get $f(x_n) = 1$ for each n, and this time the limit is one. Here, two sequences that both tend to zero are mapped by the function into sequences that converge to different limits.

It makes sense to introduce the following notion of the limit of $y = f(x)$ at a point x_0.

Definition 1.3 A number y_0 is called the *limit* of a function $y = f(x)$ as x tends to x_0 if, for any sequence $\{x_n\}$ having limit x_0 and such that $x_n \neq x_0$, the number y_0 is the limit of the sequence $\{f(x_n)\}$. This is written as

$$y_0 = \lim_{x \to x_0} f(x).$$

Here, when we refer to "any sequence" we refer to all sequences $\{x_n\}$ from the domain of $f(x)$ having the above properties.

This definition looks nice except for one thing: we must verify the desired property for *every* sequence tending to x_0. Such an infinite task is beyond the abilities of anyone. But Cauchy saved the day again by introducing another definition of function limit. Cauchy's definition turns out to be equivalent to that above, and reminds us of the definition of sequence limit.

Definition 1.4 A number y_0 is the *limit* of a function $y = f(x)$ as x tends to x_0 if for any $\varepsilon > 0$ there is a $\delta > 0$ (dependent on ε) such that for any x from the δ-neighborhood of x_0, except x_0 itself, the inequality

$$|f(x) - y_0| < \varepsilon \tag{1.1}$$

holds.

The required condition must be fulfilled for any positive ε; have we therefore simply replaced one infinite task with another? No. Our previous experience with sequence limits shows that we can solve the inequality (1.1) with

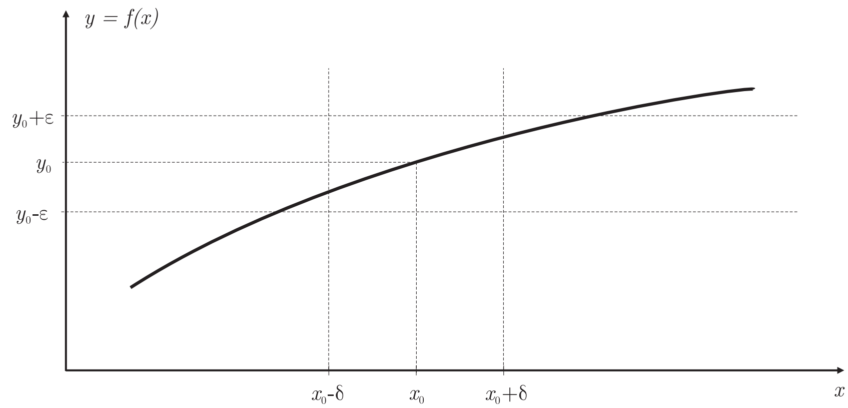

Figure 1.2 The ε–δ definition of function continuity: y_0 is the limit of $f(x)$ at x_0, if appointing a horizontal band of width 2ε about y_0, we can always find a vertical band centered at x_0 that is 2δ wide and such that all points of the graph of $y = f(x)$ (except the one corresponding to $x = x_0$) that lie in the vertical band are simultaneously in the horizontal band. This procedure should be possible for any positive ε.

respect to x for an arbitrary positive ε, and we can show that the solution domain includes the two intervals $(x_0 - \delta, x_0)$ and $(x_0, x_0 + \delta)$, where $\delta > 0$ depends on ε in some fashion.

This is presented in Figure 1.2. At point x_0, the function $f(x)$ can be defined or undefined. If it is defined and if we find that the limit at this point is equal to the defined value, that is, if

$$f(x_0) = \lim_{x \to x_0} f(x),$$

then we say that $f(x)$ is *continuous* at x_0.

We say that $f(x)$ is continuous in an open interval (a, b) if it is continuous at each point of the interval.

Continuous functions play an important role in the description of real processes and objects. Both continuous and discrete situations can be described. For example, objects are sometimes described by parameters that take on discrete values; however, we may choose to employ continuous functions as an approximation to this situation. On this idea, we base all theories of motion of material bodies — bodies that actually consist of separate atoms.

1.7 How to Measure Length

Perhaps the reader has already studied much more mathematics than we are discussing in this book. But it is probable that his or her background

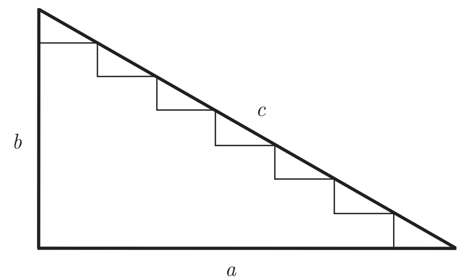

Figure 1.3 An inaccurate "approximation" of the length of a line (hypotenuse, of the length c, of a right triangle) by a polygonal curve having length $a + b$, the sum of the legs.

was slanted toward immediate applications rather than toward the theory needed to understand why mathematical "tools" work properly. We would therefore like to reconsider some topics and discuss some ideas that reside in the background of mathematics and mechanics. One important problem is this: how should we measure the length of a curve? The answer seems simple at first; we can use a tape measure or similar device. But beneath the surface of apparently simple questions we often find complex issues that require sophisticated tools for complete understanding. The calculation of length along a curve will evidently require some suitable method of approximating the desired value, followed by application of a limit passage. Here, though, even the first step — the construction of a first approximation — is not obvious. Our discussion of the problem will follow that of Henri Lebesgue (1875–1941), a mathematician known for developing a more abstract theory of integration as well as some important ideas in mechanics.

Thus, we turn to a problem of finding the length of a portion of a curve. We seek an exact result, hence ideas like placing a thread along the curve and then straightening it out against a ruler will not suffice. (It is unclear, for example, how precisely we could superimpose even a fine thread onto our curve, or what might happen to the length of the thread when we straighten it out.) So we must propose a purely geometrical mode of measurement that relies only on what we know about measuring straight segments. Because we cannot superimpose a straight segment onto a curve, our first step should be to approximate the curve by a broken (*polygonal*) line. We then try to improve the approximation again and again until the best possible approximation is reached. It is apparent that the main tool we have discussed so far — the limit — will be directly applicable to this improvement process.

We could think of many ways to approximate a curve by a polygonal line, and at first glance it seems that any polygonal line that falls close to the curve would work fine. But the following example shows that the mere requirement of closeness is not enough.

Consider the right triangle of Figure 1.3. The length of the hypotenuse c

is given by the Pythagorean theorem

$$c^2 = a^2 + b^2,$$

where a and b are the perpendicular legs. We take the hypotenuse as the curve that we wish to approximate, and form a "sawtooth" polygonal curve, as shown in the figure. The sawteeth are formed in such a way that each is similar to the original right triangle. We see that the sum of all the legs of the sawteeth equals $a + b$, independent of the number of teeth used. As the number of teeth tends to infinity, the toothed edge appears to come into coincidence with the hypotenuse of the original triangle. This gives the erroneous result $c = a + b$.

So the polygonal line must hold more tightly to the curve somehow. A better idea might be to place all of the polygonal *nodes* (i.e., corners) on the curve (and, of course, this works for a straight line). If a_k is the length of the kth polygonal segment, then the total length of the polygonal curve is the sum

$$L_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_n.$$

This looks just like the nth partial sum of a series $\sum_{k=1}^{\infty} a_n$. It seems that passage to the limit as $n \to \infty$ in L_n could solve our problem of finding the length of the curve. A difficulty lurks, however: when we summed a series, we added more and more terms *while keeping the previously added terms the same each time*. In the present case, if we increase the number of polygonal segments without bound, then the length of each will tend toward zero. This presents us with a rather strange summation of infinitely small quantities. The situation here is clearly different from that of a series, and we must decide how to deal with such sums. (We will ultimately be led to the definite integral, but we are not ready to discuss that yet.) In addition, we are faced with the idea of a curve being composed of infinitely small pieces, and we need to better understand the implications of this. The problem caught the attention of the philosopher Gottfried Leibniz (1646–1716) during his attempts to explain the structure of natural bodies. Leibniz suspected that any object is composed of infinitely small pieces. This suspicion eventually led him to the notions of differential and derivative. Leibniz actually hoped to use this idea to prove the existence of God through the use of mathematics. The attempt was brave, but from our modern viewpoint, mathematics can only demonstrate the consequences of axioms — not the axioms themselves! But it frequently happens that the errors of great men are so great that they are converted into correct assertions many generations later, as was the case with Hooke's corpuscular theory of light.

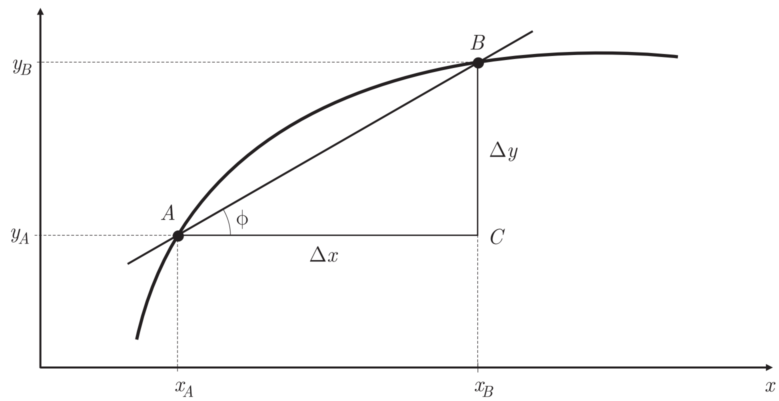

Figure 1.4 Calculation of the tangent line at point A, which is approximated by secant AB constituting an angle ϕ with the positive direction of the x-axis. Note that $\tan\phi = \Delta y/\Delta x$.

Tangent to a Curve

A particular but important problem solved by Leibniz was as follows. Given a curve described by an equation $y = f(x)$, we seek the straight line tangent to the curve at a specified point. Let us first try to reconstruct Leibniz's way of thinking about how such a tangent line should be defined.

On an intuitive basis, it is easy to see what must be done. We must find the straight line that is "closest" to the given curve at the given point. But how should "closeness" be defined in this situation?

- We know what it means for a line to be tangent to a circle: the line must have a unique point in common with the circle. But this definition does not give us a practical way to find the tangent; furthermore, we can easily imagine a curve whose tangent line meets the curve at more than one point. For example, the horizontal line $y = 1$ is tangent to the curve $y = \cos x$ at any $x = 2\pi n$.
- There is a theorem stating that any tangent to a circle is perpendicular to the radius of the circle at the point of tangency. We might be tempted to use this property as a general definition of the tangent line. But where could we take this right angle at a point of an arbitrary curve?
- The tangent to a circle can be imagined as falling at the limit position of a secant line through the desired point of tangency, when the other end of the secant tends to the desired point of tangency. Could this same idea work for a general curve?

Leibniz made the third idea the centerpiece for his definition of the tangent to an arbitrary curve. Following his lead, let us construct a tangent at point A of the curve $y = f(x)$, shown in Figure 1.4. Consider a secant line AB intersecting the curve at point B. This secant and the straight lines

AC and BC running parallel to the coordinate axes determine a triangle ACB. We denote the legs of this triangle by Δx and Δy, as shown, where $\Delta x = x_B - x_A$ is called the *increment* of the argument, and $\Delta y = y_B - y_A$ is the corresponding increment of the function. It is seen that

$$x_B = x_A + \Delta x, \qquad \Delta y = y_B - y_A = f(x_A + \Delta x) - f(x_A),$$

and that the fraction $\Delta y / \Delta x$ is equal to the tangent of the angle whose vertex is A.

Now suppose that B moves down the curve and approaches coincidence with A. We imagine the secant line rotating about A and approaching a limit state. To take this idea further, we must have a function on which we can actually perform the corresponding limit passage. For this, it makes sense to choose the tangent of the angle that the secant makes with the x-direction: this is well defined at each point B that is close to A (but distinct from A), and its limiting value is what we are seeking. Let ϕ_0 be the angle of the tangent at point A of the curve. We have

$$\tan \phi_0 = \lim_{\Delta x \to 0} \frac{f(x_A + \Delta x) - f(x_A)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \,. \qquad (1.2)$$

If this limit exists, then we can determine the desired tangent line.

Later, we shall come to understand the importance of (1.2). But first, we will see how the same formula arises in another problem.

Velocity of Motion

Sir Isaac Newton was a scientific giant, an expert in mathematics, physics, economics, and even alchemy. (It is possible that Newton spent more time on the latter than on the sciences for which he is considered a pioneer. His notes indicate a persistent search for the *philosopher's stone* — a stone that would allow one to change various other materials into gold.) Newton tackled the significant problem of finding the velocity of a point in nonuniform motion. We know that velocity in uniform motion is obtained by dividing the distance traveled by the time it takes to travel that distance. But for nonuniform motion, this gives only an *average* velocity. The average is not, for example, the value we see on a car speedometer as we are driving. There were no cars in Newton's time, but there were ships; people were also interested in calculating the motions of the stars and planets. Newton came up with an idea similar to that of Leibniz, as discussed above: instead of trying to calculate the velocity in one step, we seek to approximate it more and more closely via a limiting process. So we shall consider the distance traveled by a point over shorter and shorter time intervals. Each time, we calculate the average velocity as the ratio of distance traveled to time of travel. The limiting value of these average velocities, as the time interval tends to zero, is the *instantaneous velocity* we seek. CONTINUED...