THEORY SUPPLEMENT

A THE UNDERPINNINGS OF CALCULUS

Mathematics in its applied form has existed from time immemorial. Commercial arithmetic and the geometry of land-surveying and building construction were well-developed by 1500 B.C. Gradually people realized that simple mathematical facts may be interrelated in non-obvious ways, and that the interrelationships themselves were worthy of study. Thales (640-546 B.C.) is said to have *proved* that the sum of the angles of a triangle is two right angles. This is the oldest indication we have of the idea of proof in geometry. In the next section we look at an example of a proof.

Over the next few centuries, people began to think of geometry as being not about points, lines, and circles drawn with chalk on a slate, but about abstract entities: points tinier than the smallest speck, lines that are perfectly straight, and circles that are perfectly round. In other words, what one draws on a slate or carves into stone is merely an imperfect model of the abstract reality. Plato (427-347 B.C.) extended this view. He believed that the entire world of experience was an imperfect shadow of the true reality. However, one cannot reason about abstract things without taking some of their properties for granted. In mathematics, these assumptions are called *axioms* or *postulates*.

About 300 B.C. Euclid wrote a textbook, *The Elements*, covering a good deal of geometry, some number theory, and some more advanced topics concerning irrational numbers. It was and is the most successful textbook ever written. (It is still in print.) Euclid begins his treatment of geometry by stating several axioms concerning lines and circles. For example:

If A and B are two points, there is a circle having center A that passes through B.

This is surely a reasonable property to ascribe to the abstract points and circles that we imagine. He goes on to prove many facts about figures in the plane and in space, including the celebrated Pythagorean theorem: The area of the square drawn on the hypotenuse of a right triangle is equal to the sum of the areas of the squares drawn on the two legs.

For many years *The Elements* was revered as the pinnacle of logical reasoning. It is quite rightly regarded as a masterpiece, but its reasoning is no longer thought to be airtight. In fact, there is an error (by modern standards) in the proof of the very first proposition. The argument goes as follows.

> Starting from two points A and B, consider the circle with center A passing through B and the circle with center B passing through A. (See Figure A.1.) Let these circles cut one another at C...

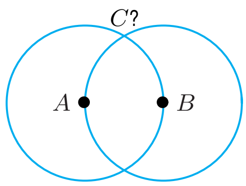

Figure A.1

But why must there be a point at which these two circles intersect? It is clear from the figure that they do, but the figure is drawn in the real world, not the abstract world of pure geometry. Perhaps when the circles we have drawn are replaced by their abstract representations, it might turn out that there isn't any point where C ought to be. Here and in several other places Euclid seems to have relied on a figure. Perhaps he did not know how to describe clearly those properties of drawings that he believed carried over to abstract geometry and guaranteed the existence of C, and so he left it to his readers to decide whether they believed that C would exist in the ideal realm. Although these deficiencies were noted in classical times, *The Elements* retained its status as the ultimate example of mathematical rigor until well into the nineteenth century. Finally, after centuries of study by many mathematicians, Hilbert (1863-1943) gave what is regarded today as the definitive treatment of Euclidean geometry. It is important to realize that none of the theorems stated by Euclid have been found to be wrong in substance; the difficulties lie entirely in Euclid's incomplete statement of the axioms on which he was relying.

Calculus belongs to a different branch of mathematics than geometry. Instead of lines, circles, and angles, calculus studies the behavior of numerical functions: specifically, functions that represent a rate of change. Calculus also provides a language for expressing laws of nature that govern everything from the behavior of the atomic nucleus to the life cycles of stars.

Some anticipations of calculus can be seen in Euclid and other classical writers, but most of the ideas appear first in the seventeenth century. Newton (1642-1727) and Leibniz (1646-1716) are generally credited with shaping the subject into a coherent theory. Newton's most famous work, *Philosophiae Naturalis Principia Mathematica* (in three volumes) appeared in 1686-1687. Its best known result is that the Laws of Planetary Motion, which had been announced by Kepler (1571-1630) on purely empirical evidence, can be deduced from simpler universal laws, such as the Law of Gravity. In addition, Newton's theory explained other astronomical phenomena, such as the irregularities in the motion of the moon, and terrestrial phenomena, such as the tides. The real significance of *Principia* lies in its demonstration that very complicated physical systems can be successfully modeled by pure mathematics. Although *Principia* uses geometrical arguments, not calculus, the ideas in it were, by Newton's own statement, generated with the aid of calculus.

After its start in the seventeenth century, calculus went for over a century without a proper axiomatic foundation. Newton wrote that it could be rigorously founded on the idea of *limits*, but he never presented his ideas in detail. A limit is, roughly speaking, the value approached by a function near a given point. During the eighteenth century many mathematicians based their work on limits, but their definition of limit was not clear. In 1784, Lagrange (1736-1813) at the Berlin Academy proposed a prize for a successful axiomatic foundation for calculus. He and others were interested in being as certain of the internal consistency of calculus as they were about algebra and geometry. No one was able to successfully respond to the challenge. It remained for Cauchy (1789-1857) to show, around 1820, that limits can be defined rigorously by means of inequalities. The modern definition of the limit, given on page 11, is essentially due to Cauchy.[1]

This rigorous definition of the limit was the advance that was needed in order to begin the axiomatic foundations of calculus, where every result is carefully proved from axioms, or from theorems that have already been proved. Courses in *analysis* follow this chain of logical reasoning.

In the textbook we concentrated on developing the solid intuitive understanding on which the rigorous approach depends, emphasizing plausibility arguments, not proofs. In this supplement we give some glimpses of the theoretical underpinnings of calculus. We hope that this brief excursion into a more theoretical world encourages you to investigate further.

B A CASE STUDY IN RIGOROUS ARGUMENT: THE BINOMIAL THEOREM

In everyday life we are often content to believe things simply by observing that they seem to be true. In mathematics, however, we decide what is true by means of logical arguments. Mathematicians attempt to eliminate all possible sources of disagreement by carefully stating *axioms* (assumptions) and *definitions*, formulating precise *theorems* (statements to be proved), and using strict rules of logic. In this section we illustrate how theorems are formulated and proved by studying the example of the binomial theorem. In the next section we see how, and why, an axiom is introduced, using as an example the completeness of the real numbers.

Recall the algebraic formulas for squaring and cubing $x+y$:

$$(x+y)^2 = x^2 + 2xy + y^2$$
$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3.$$

[1] Grabiner, Judith V. "Who Gave You the Epsilon? Cauchy and the Origins of Rigorous Calculus." *American Mathematical Monthly* 90 (1983) pp. 185-194.

We want to find a general formula for $(x+y)^n$ for any positive integer n. We do this in three stages:

- Find a pattern.
- Formulate a statement, called a *conjecture*, describing the pattern.
- Prove the conjecture.

Once we have proved the statement, it becomes a *theorem*.

Finding a pattern

First we look at some more examples. Multiplying out $(x+y)^n$ for $n = 4, 5, 6$ gives:

$$\begin{aligned}
(x+y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\
(x+y)^5 &= x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5 \\
(x+y)^6 &= x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6.
\end{aligned}$$

Notice that the exponents of x and y in each term on the right always add up to n. The reason for this is that in the expansion of

$$(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n \text{ times}},$$

each term comes from choosing x's from some of the factors and y's from the others. The total number of x's and y's chosen equals the total number of $(x+y)$'s, which is n. For example, in the expansion of $(x+y)^3$, choosing x from one of the factors and y from the other two yields a term xy^2. There are three different ways of doing this (depending on which factor x is chosen from), so there are three terms of this form, giving $3xy^2$.

We arrange the coefficients in the expansion of $(x+y)^n$ in a triangle called Pascal's triangle, after the French mathematician Blaise Pascal.

$$\begin{array}{ccccccccccccc}
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

The second row in this triangle gives the coefficients in the expansion of $(x+y)^2 = x^2 + 2xy + y^2$, namely 1, 2, and 1. The next row gives the coefficients for $(x+y)^3$, and so on. The top row gives the coefficients for the expansion $(x+y)^1 = x+y$.

There appears to be a pattern to the triangle: The outside entries are all 1s; each inside entry is equal to the sum of the entries immediately to its left and right in the row above. For example, each 10 in the fourth row has a 4 and a 6 immediately above it, and $10 = 4 + 6$.

Formulating the theorem

We want to prove that, for any n, the coefficients in the expansion of $(x+y)^n$ satisfy the pattern we have observed for $n = 1, \ldots, 6$. The general case is made easier by writing C_k^n for the coefficient of $x^{n-k}y^k$ in the expansion of $(x+y)^n$, so

$$(x+y)^n = C_0^n x^n + C_1^n x^{n-1}y + C_2^n x^{n-2}y^2 + \cdots + C_{n-1}^n xy^{n-1} + C_n^n y^n.$$

Thus, for example, $C_3^5 = 10$ because the x^2y^3 term in the expansion of $(x+y)^5$ is $10x^2y^3$.

Now

$$C_0^n \quad C_1^n \quad C_2^n \ldots \quad C_{n-1}^n \quad C_n^n$$

is the n-th row in Pascal's triangle. There are two rules that described the pattern we have observed: first, the outside entries are all 1s; second, each inside entry is the sum of the two above it. Thus, we must show, first, that $C_0^n = 1$ and $C_n^n = 1$ for all n, and, second, that

$$C_k^n = C_{k-1}^{n-1} + C_k^{n-1}, \quad 0 < k < n.$$

Notice, if $0 < k < n$, then C_k^n is an inside entry in the triangle, and C_{k-1}^{n-1} and C_k^{n-1} are the entries immediately above it. Now we can state the theorem we want to prove as follows:

The Binomial Theorem

If n is a positive integer and we write

$$(x+y)^n = C_0^n x^n + C_1^n x^{n-1}y + C_2^n x^{n-2}y^2 + \cdots + C_{n-1}^n xy^{n-1} + C_n^n y^n,$$

then

$$C_0^n = C_n^n = 1, \quad \text{for } n \geq 1,$$

and

$$C_k^n = C_{k-1}^{n-1} + C_k^{n-1}, \quad \text{for} \quad n \geq 2 \quad \text{and} \quad 0 < k < n.$$

Proof In the expansion of

$$(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n \text{ times}},$$

there is only one way of getting the term x^n, and that is by choosing an x from each factor. So, the coefficient of x^n is 1. By the same argument, the coefficient of y^n is also 1, so

$$C_0^n = C_n^n = 1.$$

To prove $C_k^n = C_{k-1}^{n-1} + C_k^{n-1}$, we write

$$(x+y)^n = C_0^n x^n + C_1^n x^{n-1}y + C_2^n x^{n-2}y^2 + \cdots + C_{n-1}^n xy^{n-1} + C_n^n y^n,$$

and we write

$$(x+y)^{n-1} = C_0^{n-1} x^{n-1} + C_1^{n-1} x^{n-2}y + C_2^{n-1} x^{n-3}y^2 + \cdots + C_{n-2}^{n-1} xy^{n-2} + C_{n-1}^{n-1} y^{n-1}.$$

Now, we will use the fact that

$$(x+y)^n = (x+y)(x+y)^{n-1}.$$

Substituting in the expressions for $(x+y)^{n-1}$ and $(x+y)^n$ gives

$$\begin{aligned}
&C_0^n x^n + C_1^n x^{n-1}y + \quad \cdots \quad + C_{n-1}^n xy^{n-1} + C_n^n y^n \\
&= (x+y)\left(C_0^{n-1}x^{n-1} + C_1^{n-1}x^{n-2}y + \quad \cdots \quad + C_{n-2}^{n-1}xy^{n-2} + C_{n-1}^{n-1}y^{n-1}\right) \\
&= x\left(C_0^{n-1}x^{n-1} + C_1^{n-1}x^{n-2}y + \quad \cdots \quad + C_{n-2}^{n-1}xy^{n-2} + C_{n-1}^{n-1}y^{n-1}\right) \\
&\qquad + y\left(C_0^{n-1}x^{n-1} + C_1^{n-1}x^{n-2}y + \quad \cdots \quad + C_{n-2}^{n-1}xy^{n-2} + C_{n-1}^{n-1}y^{n-1}\right) \\
&= C_0^{n-1}x^n + \left(C_1^{n-1} + C_0^{n-1}\right)x^{n-1}y + \cdots + \left(C_{n-1}^{n-1} + C_{n-2}^{n-1}\right)xy^{n-1} + C_{n-1}^{n-1}y^n.
\end{aligned}$$

The inside terms in this expression have the form $(C_k^{n-1} + C_{k-1}^{n-1})x^{n-k}y^k$, for $k = 1, \ldots, n-1$. The corresponding term in the expansion of $(x+y)^n$ is $C_k^n x^{n-k}y^k$. Since the expressions are equal, the coefficients of like terms must be equal, so

$$C_k^n = C_k^{n-1} + C_{k-1}^{n-1} = C_{k-1}^{n-1} + C_k^{n-1},$$

which is what we wanted to show.

A Formula for the Binomial Coefficients

The numbers C_k^n are called *binomial coefficients*. They are usually computed using the following formula, rather than by writing out Pascal's triangle. (Note that $k! = k(k-1)\cdots 3\cdot 2\cdot 1$.)

$$\boxed{C_k^n = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}}$$

This formula holds for $k = 0$ and $k = n$ if we adopt the convention that $0! = 1$:

$$C_0^n = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1 \qquad \text{and} \qquad C_n^n = \frac{n!}{n!(n-n)!} = \frac{n!}{n!} = 1.$$

To prove the formula in general, we use an important technique called *mathematical induction*. There are two steps:

- Prove the formula in the case $n = 1$.
- Prove that if it holds for a specific positive integer n then it holds for $n + 1$.

The second step, called the induction step, enables us to deduce that the formula is true for all n, as follows. Since we know by the first step that it is true for $n = 1$, by the induction step it is true for $n = 2$. But then, by the induction step again, it is true for $n = 3$, and so on.

Proof We have already proved that the formula holds for $n = 1$ since the only binomial coefficients in that case are C_0^1 and C_1^1.

Now we prove the induction step. Suppose that our formula is true for n. That is, suppose that

$$C_k^n = \frac{n!}{k!(n-k)!}, \qquad 0 \le k \le n.$$

We want to deduce the formula for $n + 1$. That is, we want to show that

$$C_k^{n+1} = \frac{(n+1)!}{k!(n+1-k)!}, \qquad 0 \le k \le n+1.$$

We already know that this is true if $k = 0$ or $k = n + 1$. If $0 < k < n + 1$, then, using the Binomial Theorem,

$$\begin{aligned}
C_k^{n+1} = C_k^n + C_{k-1}^n &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\
&= \frac{n!}{k(k-1)!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{1}{k} + \frac{1}{n-k+1}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{n-k+1+k}{k(n-k+1)}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\,\frac{(n+1)}{k(n-k+1)} = \frac{(n+1)!}{k!(n-k+1)!},
\end{aligned}$$

which is what we wanted to prove.

Problems for Section B

1. The formula for the binomial coefficients gives C_k^n as a ratio of integers. Is C_k^n necessarily an integer? Could C_k^n be a fraction? Justify your answer.
2. Look at the entries in the first few rows of Pascal's triangle on page 4. You should see a pattern of symmetry.
   (a) Describe the pattern in words.
   (b) Formulate a conjecture about the binomial coefficients C_k^n that describes the pattern mathematically.
   (c) Prove your conjecture.
3. Add the entries across the rows in Pascal's triangle for the first six rows. You should notice a pattern to the sequence of numbers that you obtain.
   (a) Formulate a general conjecture that describes the pattern.
   (b) Prove your conjecture.

C COMPLETENESS OF THE REAL NUMBERS

If two people argue about something long enough, they may eventually reveal the hidden assumptions that are the root of the argument. In the same way, mathematicians arrive at axioms by a process which is somewhat like arguing with themselves. In attempting to understand something, they question every seemingly obvious statement, hoping to eventually arrive at fundamental axioms.

We apply this method to the process of finding a root of a polynomial by zooming in on its graph. This will lead us to a subtle property of the real numbers, called *completeness*. Many proofs involving limits depend on this property.

Case Study: Finding the Roots of a Polynomial

Consider the polynomial $f(x) = 3x^3 - x^2 + 2x - 1$ on the interval $[0, 1]$. Since $f(0) = -1$ and $f(1) = 3$, we expect that the graph of f crosses the x-axis at some point $x = r$ between $x = 0$ and $x = 1$. Since the coordinates of this point are $(r, 0)$, we have $f(r) = 0$. Suppose we estimate this root by graphing the polynomial on a calculator or computer and zooming in. We start by knowing that

$$0 \le r \le 1.$$

By zooming in, we find $f(0.4) < 0$ and $f(0.5) > 0$, so r is trapped in a smaller interval

$$0.4 \le r \le 0.5.$$

Successive zooming in shows that

$$0.45 \le r \le 0.46$$
$$0.459 \le r \le 0.460$$
$$0.4598 \le r \le 0.4599.$$

At each stage, we divide the interval into tenths and pick any one for which f is negative at the left end point and positive at the right. (If f is 0 at any of the endpoints, we have found r and can stop.) Continuing this way we obtain a sequence of intervals, each one one-tenth the length of the previous one and each one containing r. (See Figure C.2.) Although a calculator will only give us a finite number of digits, we could in principle continue forever, generating an infinite sequence of intervals.

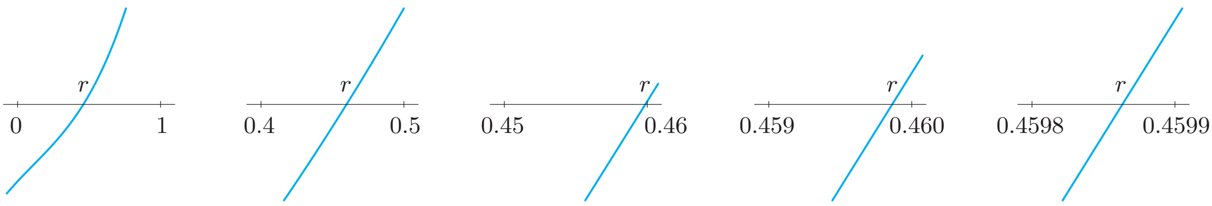

Figure C.2: Zooming in on a zero of $f(x) = 3x^3 - x^2 + 2x - 1$

It seems that this process leads us to a number r such that $f(r) = 0$. However, there are two questions that can be raised:

- How do we know[2] that this process of zooming in really does close in on a number, r?
- How do we know that $f(r) = 0$?

The Completeness Axiom

Consider the first question above (the answer to the second question is worked out in Problem 26 on page 18). The left hand endpoints of the nested intervals form an ever increasing sequence of decimals; clearly the number r that we are looking for is the smallest number that is greater than all these decimals. The completeness axiom says that, given any nonempty set of numbers, if there is any number which is greater than or equal to all of the numbers in the set, then there is a smallest such number.[3] A number which is greater than or equal to all the numbers in a set is called an *upper bound* for the set. We have the following:

The Completeness Axiom

Any nonempty set of real numbers which has an upper bound has a *least* upper bound.

Example 1 For each of the following sets, say whether it has an upper bound. If so, give the least upper bound.

(a) The set of x such that $-2 < x < 3$.
(b) The set of x such that $-2 \leq x \leq 3$.
(c) The set of all integers.
(d) The sequence $0.9,\ 0.99,\ 0.999, \ldots$.

Solution

(a) The numbers 3, 4, and π are all upper bounds; 3 is the least upper bound.
(b) Same as part (a); an upper bound for a set can be in the set, since it only has to be greater than *or equal to* each number in the set.
(c) There is no upper bound for this set; no matter how large a number we choose for the upper bound, there will always be some integer bigger than it.
(d) All the numbers are less than 1, and 1 is the smallest number with this property. Thus, this sequence has a least upper bound of 1.

In the previous example we could see directly what the least upper bounds were. In other situations it may not be obvious. The completeness axiom guarantees the existence of the least upper bound but offers no help in finding it.

The Nested Interval Theorem

Now we see how the completeness axiom ensures that the process of zooming in closes in on a specific number.

Nested Interval Theorem

Given an infinite sequence of closed intervals, $[a_n, b_n]$, each one contained within the previous one, then there is at least one number in all the intervals.

[2]This is related to the question raised on page 2 about the existence of an intersection point C between two circles.
[3]It is possible to give a definition of the real numbers in which the completeness axiom becomes a theorem.

Proof Because each interval $[a_n, b_n]$ is contained in the previous one, each a_n must be at least as large as the previous one, so

$$a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n \leq \cdots.$$

Similarly, each b_n is no larger than its predecessor, so

$$\cdots \leq b_n \leq \cdots \leq b_3 \leq b_2 \leq b_1.$$

All the a_n are bounded above by b_1, and in fact by every b_n. Therefore, by the completeness axiom, we know that the a_n have a least upper bound; call it r. Since r is an upper bound, $r \geq a_n$ for all n. Since r is the least upper bound, r must be less than or equal to each of the upper bounds b_n. Thus, r is in all the intervals.

Note that in our statement of the Nested Interval Property, we did not assume that the lengths of the intervals approached zero, so in general there may be more than one number r in all the intervals. However, when the lengths do approach zero, as in the case of finding a root by zooming in, there is a *unique* number r (see Problem 2).

The Intermediate Value Theorem

When we considered the polynomial $3x^3 - x^2 + 2x - 1$, we assumed it must have a root between $x = 0$ and $x = 1$ because it was negative at $x = 0$ and positive at $x = 1$. More generally, our intuitive notion of continuity tells us that, as we follow the graph of a continuous function f from some point $(a, f(a))$ to another point $(b, f(b))$, then f must take on all intermediate values between $f(a)$ and $f(b)$. (See Figure C.3.) This is:

Intermediate Value Theorem

Suppose f is continuous on a closed interval $[a, b]$. If k is any number between $f(a)$ and $f(b)$, then there is at least one number c in $[a, b]$ such that $f(c) = k$.

Problems 26 and 27 on page 18 suggest a way of proving the Intermediate Value Theorem using the Nested Interval Theorem.

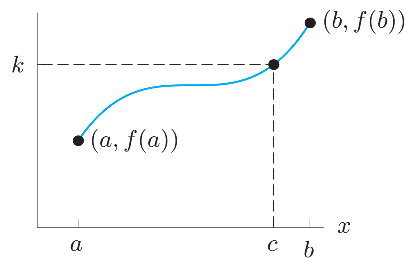

Figure C.3: The Intermediate Value Theorem

Problems for Section C

1. **(a)** Using the definitions in this section as a guide, define the following terms:
   (i) A lower bound of a set of numbers
   (ii) The greatest lower bound of a set of numbers

   (b) State the completeness axiom in terms of lower bounds.

2. Let r be a number contained in each of a sequence of nested intervals $[a_n, b_n]$. Suppose that the width of the intervals, $|b_n - a_n|$, goes to 0 as $n \to \infty$. Prove that r is unique. [Hint: Suppose there were two such numbers, and argue to a contradiction.]

3. In this problem we will use the completeness axiom to show that an infinite decimal expansion actually defines a real number, and that the first n digits of the expansion give the number accurate to n decimal places. Let x_n be the number defined by the first n digits of the expansion; we call x_n the n-th *truncation* of the expansion.

   (a) For any n, show that $x_n + (1/10^n)$ is an upper bound for the set of all the truncations.

   (b) Deduce that there is a real number, c, such that $x_n \le c \le x_n + (1/10)^n$ for all n. Thus x_n represents c accurate to n places, so it's reasonable to say that c is the number represented by the infinite decimal expansion.

D LIMITS AND CONTINUITY

In this section we use the example of limits and continuity to illustrate how formal definitions are developed from intuitive ideas.

Definition of Limit

By the beginning of the 19th century, calculus had proved its worth, and there was no doubt about the correctness of its answers. However, it was not until the work of the French mathematician Augustin Cauchy (1789–1857) that a formal definition of the limit was given, similar to the following:

> Suppose a function f, is defined on an interval around c, except perhaps not at the point $x = c$. We define the **limit** of the function $f(x)$ as x approaches c, written $\lim_{x\to c} f(x)$, to be a number L (if one exists) such that $f(x)$ is as close to L as we please whenever x is sufficiently close to c (but $x \neq c$). If L exists, we write
>
> $$\lim_{x\to c} f(x) = L.$$

Shortly, we will see how "as close as we please" and "sufficiently close" can be given a precise meaning using inequalities. First, we look at $\lim\limits_{\theta\to 0}(\sin\theta/\theta)$ more closely (see Example 1 on page 68 of the textbook).

Example 1 By graphing $y = (\sin\theta)/\theta$ in an appropriate window, find how close θ should be to 0 in order to make $(\sin\theta)/\theta$ within 0.01 of 1.

Solution Since we want $(\sin\theta)/\theta$ to be within 0.01 of 1, we set the y-range on the graphing window to go from 0.99 to 1.01. Our first attempt with $-0.5 \le \theta \le 0.5$ yields the graph in Figure D.4. Since we want the y-values to stay within the range $0.99 < y < 1.01$, we do not want the graph to leave the window through the top or bottom. By trial and error, we find that changing the θ-range to $-0.2 \le \theta \le 0.2$ gives the graph in Figure D.5. Thus, the graph suggests that $(\sin\theta)/\theta$ will be within 0.01 of 1 whenever θ is within 0.2 of 0. Proving this requires an analytical argument, not just graphs from a calculator.

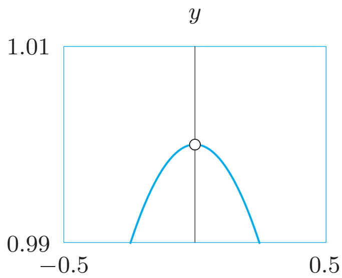

Figure D.4: $(\sin\theta)/\theta$ with $-0.5 \le \theta \le 0.5$

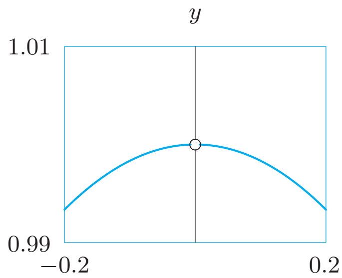

Figure D.5: $(\sin\theta)/\theta$ with $-0.2 \le \theta \le 0.2$

When we say "$f(x)$ is as close to L as we please," we mean that we can specify a maximum distance between $f(x)$ and L. We express the distance using absolute values:

$$|f(x) - L| = \text{Distance between } f(x) \text{ and } L.$$

Using ϵ (the Greek letter epsilon) to stand for the distance we have specified, we write

$$|f(x) - L| < \epsilon$$

to indicate that the maximum distance between $f(x)$ and L is less than ϵ. In Example 2 we used $\epsilon = 0.01$. In a similar manner we interpret "x is sufficiently close to c" as specifying a maximum distance between x and c:

$$|x - c| < \delta,$$

where δ (the Greek letter delta) tells us how close x should be to c. In Example 2 we found $\delta = 0.2$.

If $\lim_{x\to c} f(x) = L$, we know that no matter how narrow the band determined by ϵ in Figure D.6, there's always a δ which makes the graph stay within the band for $c - \delta < x < c + \delta$.

Thus we restate the definition of a limit, using symbols:

Definition of Limit

We define $\lim_{x\to c} f(x)$ to be the number L (if one exists) such that for any $\epsilon > 0$ (as small as we want), there is a $\delta > 0$ (sufficiently small) such that if $|x - c| < \delta$ and $x \ne c$, then $|f(x) - L| < \epsilon$.

Realize that the point of this definition is that for any ϵ we are given, we need to be able to determine a corresponding δ. One way to do this is to give an explicit expression for δ in terms of ϵ.

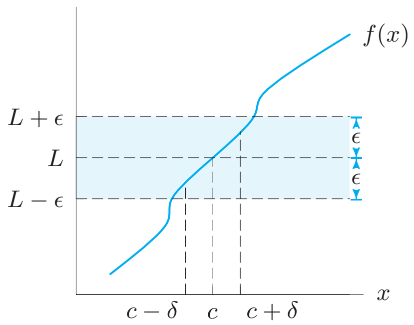

Figure D.6: What the definition of the limit means practically

Example 2 Use algebra to find a maximum distance between x and 2 which ensures that x^2 is within 0.1 of 4. Use a similar argument to show that $\lim_{x \to 2} x^2 = 4$.

Solution We write $x = 2 + h$. We want to find the values of h making x^2 within 0.1 of 4. We know

$$x^2 = (2+h)^2 = 4 + 4h + h^2,$$

so x^2 differs from 4 by $4h + h^2$. Since we want x^2 to be within 0.1 of 4, we need

$$\left|x^2 - 4\right| = \left|4h + h^2\right| = |h| \cdot |4 + h| < 0.1.$$

Assuming $0 < |h| < 1$, we know $|4 + h| < 5$, so we need

$$\left|x^2 - 4\right| < 5\,|h| < 0.1.$$

Thus, if we choose h such that $0 < |h| < 0.1/5 = 0.02$, then x^2 is less than 0.1 from 4.

An analogous argument using any small ϵ instead of 0.1 shows that if we take $\delta = \epsilon/5$, then

$$\left|x^2 - 4\right| < \epsilon \quad \text{for all} \quad |x - 2| < \epsilon/5.$$

Thus, we have used the definition to show that

$$\lim_{x \to 2} x^2 = 4.$$

It is important to understand that the ϵ, δ definition by itself does not make it easier to calculate limits. The advantage of the ϵ, δ definition is that it makes it possible to put calculus on a rigorous foundation. From this foundation, we can prove the following properties. See Problems 13–16.

Theorem: Properties of Limits

Assuming all the limits on the right hand side exist:

1. If b is a constant, then $\lim_{x \to c} (bf(x)) = b\left(\lim_{x \to c} f(x)\right)$.
2. $\lim_{x \to c} (f(x) + g(x)) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$.
3. $\lim_{x \to c} (f(x)g(x)) = \left(\lim_{x \to c} f(x)\right)\left(\lim_{x \to c} g(x)\right)$.
4. $\lim_{x \to c} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x \to c} f(x)}{\lim_{x \to c} g(x)}$, provided $\lim_{x \to c} g(x) \neq 0$.
5. For any constant k, $\lim_{x \to c} k = k$.
6. $\lim_{x \to c} x = c$.

These properties underlie many limit calculations, though we seldom acknowledge them explicitly.

Example 3 Explain how the limit properties are used in the following calculation:

$$\lim_{x \to 3} \frac{x^2 + 5x}{x + 9} = \frac{3^2 + (5)(3)}{3 + 9} = 2.$$

Solution

We calculate this limit in stages, using the limit properties to justify each step:

$$\begin{aligned}
\lim_{x\to 3}\frac{x^2+5x}{x+9} &= \frac{\lim\limits_{x\to 3}(x^2+5x)}{\lim\limits_{x\to 3}(x+9)} && \text{Property 4 (since } \lim_{x\to 3}(x+9)\neq 0)\\
&= \frac{\lim\limits_{x\to 3}(x^2)+\lim\limits_{x\to 3}(5x)}{\lim\limits_{x\to 3}x+\lim\limits_{x\to 3}9} && \text{Property 2}\\
&= \frac{\left(\lim\limits_{x\to 3}x\right)^2+5\left(\lim\limits_{x\to 3}x\right)}{\lim\limits_{x\to 3}x+\lim\limits_{x\to 3}9} && \text{Properties 1 and 3}\\
&= \frac{3^2+(5)(3)}{3+9}=2. && \text{Properties 5 and 6}
\end{aligned}$$

One- and Two-Sided Limits

When we write

$$\lim_{x\to 2} f(x),$$

we mean the number that $f(x)$ approaches as x approaches 2 *from both sides*. We examine values of $f(x)$ as x approaches 2 through values greater than 2 (such as 2.1, 2.01, 2.003) and values less than 2 (such as 1.9, 1.99, 1.994). If we want x to approach 2 only through values greater than 2, we write

$$\lim_{x\to 2^+} f(x)$$

for the number that $f(x)$ approaches (assuming such a number exists). Similarly,

$$\lim_{x\to 2^-} f(x)$$

denotes the number (if it exists) obtained by letting x approach 2 through values less than 2. We call $\lim\limits_{x\to 2^+} f(x)$ a *right-hand limit* and $\lim\limits_{x\to 2^-} f(x)$ a *left-hand limit*.

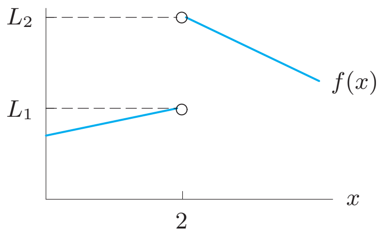

Figure D.7: Left- and right-hand limits at $x = 2$

For the function graphed in Figure D.7, we have

$$\lim_{x\to 2^-} f(x) = L_1 \qquad \lim_{x\to 2^+} f(x) = L_2.$$

If the left- and right-hand limits were equal, that is, if $L_1 = L_2$, then it could easily be proved that $\lim\limits_{x\to 2} f(x)$ exists and $\lim\limits_{x\to 2} f(x) = L_1 = L_2$. Since $L_1 \neq L_2$ in Figure D.7, we see that $\lim\limits_{x\to 2} f(x)$ does not exist in this case.

When Limits Do Not Exist

Whenever there is not a number L such that $\lim\limits_{x\to c} f(x) = L$, we say $\lim\limits_{x\to c} f(x)$ does not exist. In addition to cases in which the left- and right-hand limits are not equal, there are some other cases in which limits fail to exist. Here are three examples.

Example 4 Explain why $\lim_{x\to 2} \frac{|x-2|}{x-2}$ doesn't exist.

Solution Figure D.8 shows the problem: The right-hand limit and the left-hand limit are different. For $x > 2$, we have $|x-2| = x-2$, so as x approaches 2 from the right,

$$\lim_{x\to 2^+} \frac{|x-2|}{x-2} = \lim_{x\to 2^+} \frac{x-2}{x-2} = \lim_{x\to 2^+} 1 = 1.$$

Similarly, if $x < 2$, then $|x-2| = 2-x$ so

$$\lim_{x\to 2^-} \frac{|x-2|}{x-2} = \lim_{x\to 2^-} \frac{2-x}{x-2} = \lim_{x\to 2^-} (-1) = -1.$$

So if $\lim_{x\to 2} \frac{|x-2|}{x-2} = L$ then L would have to be both 1 and -1. Since L cannot have two different values, the limit does not exist.

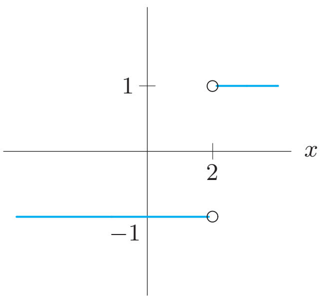

Figure D.8: Graph of $\frac{|x-2|}{x-2}$

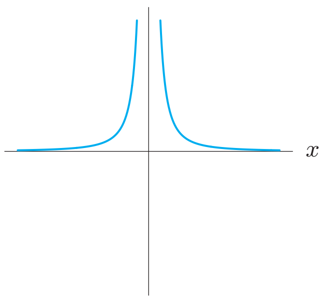

Figure D.9: Graph of $\frac{1}{x^2}$

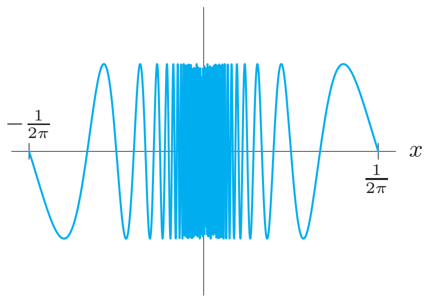

Figure D.10: Graph of $\sin\left(\frac{1}{x}\right)$

Example 5 Explain why $\lim_{x\to 0} \frac{1}{x^2}$ doesn't exist.

Solution As x approaches zero, $1/x^2$ becomes arbitrarily large, so it can't stay close to any finite number L. See Figure D.9. Therefore we say $1/x^2$ has no limit as $x \to 0$.

Example 6 Explain why $\lim_{x\to 0} \sin\left(\frac{1}{x}\right)$ doesn't exist.

Solution We know that the sine function has values between -1 and 1. The graph in Figure D.10 oscillates more and more rapidly as $x \to 0$. There are x-values as close to 0 as we like where $\sin(1/x) = 0$. There are also x-values as close to 0 as we like where $\sin(1/x) = 1$. So if the limit existed, it would have to be both 0 and 1. Thus, the limit does not exist.

Limits at Infinity

Sometimes we want to know what happens to $f(x)$ as x gets large, that is, the end behavior of f.

If $f(x)$ gets as close to a number L as we please when x gets sufficiently large, then we write

$$\lim_{x\to\infty} f(x) = L.$$

Similarly, if $f(x)$ approaches L as x gets more and more negative, then we write

$$\lim_{x\to-\infty} f(x) = L.$$

The symbol ∞ does not represent a number. Writing $x \to \infty$ means that we consider arbitrarily large values of x. If the limit of $f(x)$ as $x \to \infty$ or $x \to -\infty$ is L, we say that the graph of f has a *horizontal asymptote* $y = L$.

Example 7 Investigate $\displaystyle\lim_{x\to\infty} \frac{1}{x}$ and $\displaystyle\lim_{x\to-\infty} \frac{1}{x}$.

Solution A graph of $f(x) = \frac{1}{x}$ in a large window shows $1/x$ approaching zero as x increases in either the positive or the negative direction (See Figure D.11). This is as we would expect, since dividing 1 by larger and larger numbers yields answers which are smaller and smaller. This suggests that

$$\lim_{x\to\infty} \frac{1}{x} = \lim_{x\to-\infty} \frac{1}{x} = 0,$$

and $f(x) = 1/x$ has $y = 0$ as a horizontal asymptote as $x \to \pm\infty$.

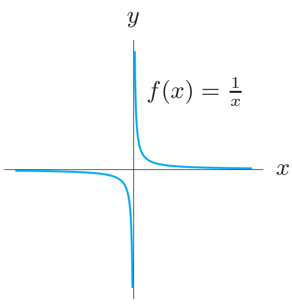

Figure D.11: The end behavior of $f(x) = \frac{1}{x}$

Definition of Continuity

We can now define continuity. Recall that the idea of continuity rules out breaks, jumps, or holes by demanding that the behavior of a function *near* a point be consistent with its behavior *at* the point:

The function f is **continuous** at $x = c$ if f is defined at $x = c$ and

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

In other words, $f(x)$ is as close as we please to $f(c)$ provided x is close enough to c. The function is **continuous on an interval** $[a, b]$ if it is continuous at every point in the interval.[4]

Constant functions and $f(x) = x$ are continuous. (See Problem 16.) Using the continuity of sums and products, we can show that any polynomial is continuous. Proving that $\sin x$, $\cos x$, and e^x are continuous is more difficult. The following theorem, based on the properties of limits on page 12, makes it easier to decide whether a given function is continuous.

Theorem: Continuity of Sums, Products, and Quotients of Functions

Suppose that f and g are continuous on an interval and that b is a constant. Then, on that same interval,

1. $bf(x)$ is continuous
2. $f(x) + g(x)$ is continuous
3. $f(x)g(x)$ is continuous
4. $f(x)/g(x)$ is continuous, provided $g(x) \neq 0$ on the interval.

We prove the first of these properties.

Proof To prove that $bf(x)$ is continuous, pick any point c in the interval. We must show that $\lim_{x\to c} bf(x) = bf(c)$. Since $f(x)$ is continuous, we already know that $\lim_{x\to c} f(x) = f(c)$. So, by the first property of limits,

$$\lim_{x\to c}(bf(x)) = b\left(\lim_{x\to c} f(x)\right) = bf(c).$$

Since c was chosen arbitrarily, we have shown that $bf(x)$ is continuous at every point in the interval.

Theorem: Continuity of Composite Functions

If f and g are continuous, and if the composite function $f(g(x))$ is defined on an interval, then $f(g(x))$ is continuous on that interval.

Assuming the continuity of $\sin x$ and e^x, this result shows us, for example, that $\sin(e^x)$ and $e^{\sin x}$ are both continuous. A proof of the continuity of composite functions is outlined in Problem 17.

Problems for Section D

1. Consider the function $(\sin\theta)/\theta$. Estimate how close θ should be to 0 to make $(\sin\theta)/\theta$ stay within 0.001 of 1.

2. The function $g(\theta) = (\sin\theta)/\theta$ is not defined at $\theta = 0$. Is it possible to define $g(0)$ in such a way that g is continuous at $\theta = 0$? Explain your answer.

Use a graph to estimate each of the limits in Problems 3–6.

3. $\displaystyle\lim_{\theta\to 0}\frac{\sin(2\theta)}{\theta}$ (use radians)

4. $\displaystyle\lim_{\theta\to 0}\frac{\cos\theta - 1}{\theta}$ (use radians)

5. $\displaystyle\lim_{\theta\to 0}\frac{\sin\theta}{\theta}$ (use degrees)

6. $\displaystyle\lim_{\theta\to 0}\frac{\theta}{\tan(3\theta)}$ (use radians)

7. Consider the limit

$$\lim_{x\to 0+} x^x.$$

Estimate this limit either by evaluating x^x for smaller and smaller positive values of x (say $x = 0.1, 0.01, 0.001, \ldots$) or by zooming in on the graph of $y = x^x$ near $x = 0$.

[4] If c is an endpoint of the interval, we define continuity at $x = c$ using one-sided limits at c.

8. **(a)** Give an example of a function such that $\lim_{x\to 2} f(x) = \infty$.
(b) Give an example of a function such that $\lim_{x\to 2} f(x) = -\infty$.

9. Consider the function $f(x) = \sin(1/x)$.

(a) Find a sequence of x-values that approach 0 such that $\sin(1/x) = 0$.
[Hint: Use the fact that $\sin(\pi) = \sin(2\pi) = \sin(3\pi) = \ldots = \sin(n\pi) = 0$.]
(b) Find a sequence of x-values that approach 0 such that $\sin(1/x) = 1$.
[Hint: Use the fact that $\sin(n\pi/2) = 1$ if $n = 1, 5, 9, \ldots$.]
(c) Find a sequence of x-values that approach 0 such that $\sin(1/x) = -1$.
(d) Explain why your answers to any two of parts (a)–(c) show that $\lim_{x\to 0} \sin(1/x)$ does not exist.

10. Write the definition of the following statement both in words and in symbols:

$$\lim_{h\to a} g(h) = K.$$

11. For each of the following functions do the following:

(i) Make a table of values of $f(x)$ for $x = a + 0.1$, $a + 0.01$, $a + 0.001$, $a + 0.0001$, $a - 0.1$, $a - 0.01$, $a - 0.001$, and $a - 0.0001$.
(ii) Make a conjecture about the value of $\lim_{x\to a} f(x)$.
(iii) Graph the function to see if it is consistent with your answers to parts (i) and (ii).
(iv) Find an interval for x containing a such that the difference between your conjectured limit and the value of the function is less than 0.01 on that interval. (In other words, find a window of height 0.02 such that the graph exits the sides of the window and not the top or bottom of the window.)

(a) $f(x) = \dfrac{x^2 - 4}{x - 2}, \quad a = 2$
(b) $f(x) = \dfrac{x^2 - 9}{x - 3}, \quad a = 3$
(c) $f(x) = \dfrac{\sin x - 1}{x - \pi/2}, \quad a = \dfrac{\pi}{2}$
(d) $f(x) = \dfrac{\sin 5x - 1}{x - \pi/2}, \quad a = \dfrac{\pi}{2}$
(e) $f(x) = \dfrac{e^{2x-2} - 1}{x - 1}, \quad a = 1$
(f) $f(x) = \dfrac{e^{0.5x-1} - 1}{x - 2}, \quad a = 2$

12. Assuming that limits as $x \to \infty$ have the properties listed for limits as $x \to c$ on page 12, use algebraic manipulations to evaluate $\lim_{x\to\infty}$ for the following functions:

(a) $f(x) = \dfrac{x + 3}{2 - x}$ **(b)** $f(x) = \dfrac{x^2 + 2x - 1}{3 + 3x^2}$
(c) $f(x) = \dfrac{x^2 + 4}{x + 3}$ **(d)** $f(x) = \dfrac{2x^3 - 16x^2}{4x^2 + 3x^3}$
(e) $f(x) = \dfrac{x^4 + 3x}{x^4 + 2x^5}$ **(f)** $f(x) = \dfrac{3e^x + 2}{2e^x + 3}$
(g) $f(x) = \dfrac{2e^{-x} + 3}{3e^{-x} + 2}$

13. This problem suggests a proof of the first property of limits on page 12: $\lim_{x\to c} bf(x) = b \lim_{x\to c} f(x)$.

(a) First, prove the property in the case $b = 0$.
(b) Now suppose that $b \neq 0$. Let $\epsilon > 0$. Show that if $|f(x) - L| < \epsilon/|b|$, then $|bf(x) - bL| < \epsilon$.
(c) Finally, prove that if $\lim_{x\to c} f(x) = L$ then $\lim_{x\to c} bf(x) = bL$. [Hint: Choose δ so that if $|x - c| < \delta$, then $|f(x) - L| < \epsilon/|b|$.]

14. Prove the second property of limits: $\lim_{x\to c} (f(x) + g(x)) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$. Assume that the limits on the right exist.

15. This problem suggests a proof of the third property of limits:

$$\lim_{x\to c} (f(x)g(x)) = \left(\lim_{x\to c} f(x)\right)\left(\lim_{x\to c} g(x)\right)$$

(assuming the limits on the right exist). Let $L_1 = \lim_{x\to c} f(x)$ and $L_2 = \lim_{x\to c} g(x)$.

(a) First, show that if $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$, then $\lim_{x\to c} (f(x)g(x)) = 0$.
(b) Show algebraically that $f(x)g(x) = (f(x) - L_1)(g(x) - L_2) + L_1 g(x) + L_2 f(x) - L_1 L_2$.
(c) Use the second limit property (see Problem 14) to explain why $\lim_{x\to c} (f(x) - L_1) = \lim_{x\to c} (g(x) - L_2) = 0$.
(d) Use parts (a) and (c) to explain why $\lim_{x\to c} (f(x) - L_1)(g(x) - L_2) = 0$.
(e) Finally, use parts (b) and (d) and the first and second limit properties to show that $\lim_{x\to c} (f(x)g(x)) = \left(\lim_{x\to c} f(x)\right)\left(\lim_{x\to c} g(x)\right)$.

16. Show that the following functions are both continuous everywhere.
(a) $f(x) = k$ (a constant) **(b)** $g(x) = x$

17. This problem suggests a proof of the theorem on continuity of composite functions on page 16: If f and g are continuous and the composite function $f(g(x))$ is defined on an interval, then $f(g(x))$ is continuous on that interval.

Let c be a point inside the interval where $f(g(x))$ is defined. We must show that $\lim_{x\to c} f(g(x)) = f(g(c))$. Let $d = g(c)$. Then the continuity of f at d means that $\lim_{y\to d} f(y) = f(d)$. Thus, for a given $\epsilon > 0$, we can choose $\delta > 0$ so that $|y - d| < \delta$ implies $|f(y) - f(d)| < \epsilon$.

Now, take $y = g(x)$ and show that the continuity of g means that we can find a $\delta_1 > 0$ such that, if $|x - c| < \delta_1$, then $|g(x) - d| < \delta$. Explain how this establishes the continuity of $f(g(x))$ at $x = c$.

For each value of ϵ in Problems 18–19, find a positive value of δ such that the graph of the function leaves the window $a - \delta < x < a + \delta$, $b - \epsilon < y < b + \epsilon$ by the sides and not through the top or bottom.

18. $f(x) = -2x + 3$; $a = 0$; $b = 3$; $\epsilon = 0.2, 0.1, 0.02, 0.01, 0.002, 0.001$.

19. $g(x) = -x^3 + 2$; $a = 0$; $b = 2$; $\epsilon = 0.1, 0.01, 0.001$.

20. Show that $\lim_{x\to 0}(-2x+3) = 3$. You may use the result of Problem 18.

21. Show that $\lim_{x\to 0}(-x^3 + 2) = 2$. [Hint: Try $\delta = \epsilon^{1/3}$.]

In Problems 22–24, modify the definition of limit on page 11 to give a definition of each of the following.

22. A right-hand limit

23. A left-hand limit

24. $\lim_{x\to\infty} f(x) = L$

25. Consider the function $f(x) = \begin{cases} x \sin\left(\dfrac{1}{x}\right) & x \neq 0 \\ 0 & x = 0. \end{cases}$

Show that f is continuous everywhere, but that it is neither always increasing nor always decreasing on the interval $[0, \epsilon]$ for any $\epsilon > 0$, no matter how small.

26. On page **??** we showed how to find a sequence of intervals $[a_n, b_n]$ that close in on a root r of $f(x) = 3x^3 - x^2 + 2x - 1$. In this problem we use the continuity of f to prove that r is in fact a root, that is, that $f(r) = 0$. You may assume that all the intervals have been chosen so that $f(a_n) < 0$ and $f(b_n) > 0$.

(a) Suppose $f(r) = L > 0$. Use the definition of continuity with any ϵ such that $\epsilon < L$ to choose a δ such that

$$|f(x) - L| < \epsilon \quad \text{for all} \quad |x - r| < \delta.$$

Find an a_n in the interval $[r - \delta, r + \delta]$ and arrive at a contradiction involving $f(a_n)$.

(b) Suppose $f(r) = L < 0$. Make a similar argument to arrive at a contradiction involving a b_n.

(c) Conclude that $f(r) = 0$.

27. Adapt the zooming argument on page **??** and the argument in Problem 26 to prove the Intermediate Value Theorem: If f is continuous on $[a, b]$ and k is between $f(a)$ and $f(b)$, there is a point c in $[a, b]$ with $f(c) = k$. [Hint: Consider $g(x) = f(x) - k$ and look for a zero of g.]

E DIFFERENTIABILITY AND LINEAR APPROXIMATION

In this section we analyze the tangent line approximation and its error. This leads us to a different way of looking at differentiability. Recall that:

> A function f is said to be **differentiable at** $x = a$ if $f'(a)$ exists.

Most functions we deal with have a derivative at every point in their domain; they are said to be *differentiable everywhere*.

How Can We Recognize Whether a Function Is Differentiable?

If a function has a derivative at a point, its graph must have a tangent line there; the slope of the tangent line is the derivative. When we zoom in on the graph of the function, we see a nonvertical straight line.

Occasionally we meet a function which fails to have a derivative at a few points. For example, a discontinuous function whose graph has a break at some point cannot have a derivative at that point. Some of the ways in which a function can fail to be differentiable at a point are if:

- The function is not continuous at the point.
- The graph has a sharp corner at that point.
- The graph has a vertical tangent line.

Figure E.12 shows a function which appears to be differentiable at all points except $x = a$ and $x = b$. There is no tangent at A because the graph has a corner there. As x approaches a from the left, the slope of the line joining P to A converges to some positive number. As x approaches a from the right, the slope of the line joining P to A converges to some negative number. Thus the slopes approach different numbers as we approach $x = a$ from different sides. Therefore the function is not differentiable at $x = a$. At B, there is no sharp corner, but as x approaches b, the slope of the line joining B to Q does not converge; it just keeps growing larger and larger. This reflects the fact that the graph has a vertical tangent at B. Since the slope of a vertical line is not defined, the function is not differentiable at $x = b$.

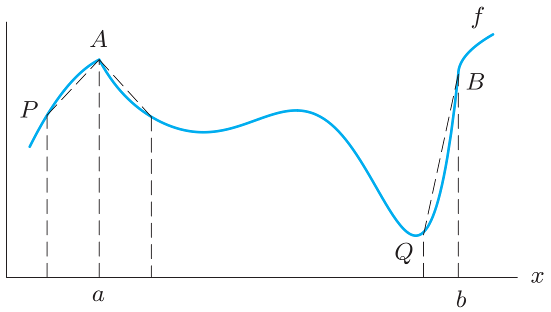

Figure E.12: A function which is not differentiable at A or B

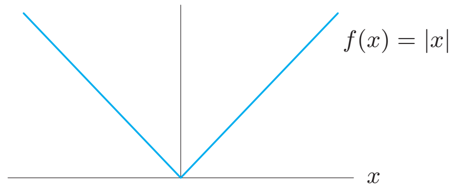

Figure E.13: Graph of absolute value function, showing point of non-differentiability at $x = 0$

Examples of Nondifferentiable Functions

The best-known function with a corner is the *absolute value* function defined as follows:

$$\boxed{f(x) = |x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}}$$

The graph of this function is in Figure E.13. Near $x = 0$, even close-up views of the graph of $f(x)$ look the same, so this is a corner which can't be straightened out by zooming in.

Example 1 Try to compute the derivative of the function $f(x) = |x|$ at $x = 0$. Is f differentiable there?

Solution To find the slope at $x = 0$, we want to look at

$$\lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} \frac{|h| - 0}{h} = \lim_{h\to 0} \frac{|h|}{h}.$$

As h approaches 0 from the right, h is always positive, so $|h| = h$, and the ratio is always 1. As h approaches 0 from the left, h is negative, so $|h| = -h$, and the ratio is -1. Since the limits are different from each side, the limit of the difference quotient does not exist. Thus, the absolute value function is not differentiable at $x = 0$. The limits of 1 and -1 correspond to the fact that the slope of the right-hand part of the graph is 1, and the slope of the left-hand part is -1.

Example 2 Investigate the differentiability of $f(x) = x^{1/3}$ at $x = 0$.

Solution This function is smooth at $x = 0$ (no sharp corners) but appears to have a vertical tangent there. (See Figure E.14.) Looking at the difference quotient at $x = 0$, we see

$$\lim_{h\to 0}\frac{(0+h)^{1/3}-0^{1/3}}{h} = \lim_{h\to 0}\frac{h^{1/3}}{h} = \lim_{h\to 0}\frac{1}{h^{2/3}}.$$

As $h \to 0$ the denominator becomes small, so the fraction grows without bound. Hence, the function fails to have a derivative at $x = 0$.

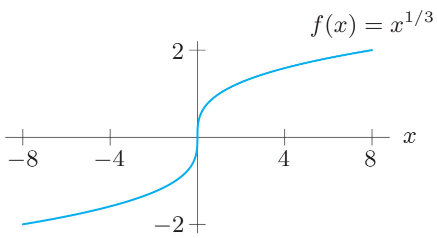

Figure E.14: Continuous function not differentiable at $x = 0$: Vertical tangent

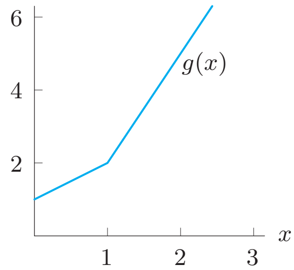

Figure E.15: Continuous function not differentiable at $x = 1$

Example 3 Consider the function given by the formulas

$$g(x) = \begin{cases} x+1 & \text{if} \quad x \le 1 \\ 3x-1 & \text{if} \quad x > 1. \end{cases}$$

This kind of function is called *piecewise linear* because each part of it is linear. Draw the graph of g. Is g continuous? Is g differentiable at $x = 1$?

Solution The graph in Figure E.15 has no breaks in it, so the function is continuous. However, the graph has a corner at $x = 1$ which no amount of magnification will remove. To the left of $x = 1$, the slope is 1; to the right of $x = 1$, the slope is 3. Thus, the difference quotient at $x = 1$ will fail to have a limit, and so the function g is not differentiable at $x = 1$.

A great deal of interest has been sparked in the last few years in the study of curves which do not possess derivatives *anywhere*. These curves, known as *fractals*, arise in the modeling of natural processes that are random and chaotic, such as the path of a water molecule in a glass of water. As the molecule bounces off of its neighbors haphazardly, it traces out a path with many jagged, nondifferentiable corners. Although the path may be smooth between collisions, it can be modeled effectively by a curve that is not differentiable anywhere. The coastlines of Maine and Washington are also examples. They never straighten out, no matter how close we look.

Differentiability and Linear Approximation

When we zoom in on the graph of a differentiable function, it looks like a straight line. In fact, the graph is not exactly a straight line when we zoom in; however, its deviation from straightness is so small that it can't be detected by the naked eye. Let's examine what this means. The straight line that we think we see when we zoom in on the graph of $f(x)$ at $x = a$ has slope equal to the derivative, $f'(a)$, so the equation is

$$y = f(a) + f'(a)(x - a).$$

The fact that the graph looks like a line means that y is a good approximation to $f(x)$. (See Figure E.16.) This suggests the following definition:

The Tangent Line Approximation

Suppose f is differentiable at a. Then, for values of x near a, the tangent line approximation to $f(x)$ is

$$f(x) \approx f(a) + f'(a)(x-a).$$

The expression $f(a) + f'(a)(x-a)$ is called the *local linearization* of f near $x = a$. We are thinking of a as fixed, so that $f(a)$ and $f'(a)$ are constant.

The **error**, $E(x)$, in the approximation is defined by

$$E(x) = f(x) - f(a) - f'(a)(x-a).$$

It can be shown that the tangent line approximation is the best linear approximation to f near a. See Problem 15.

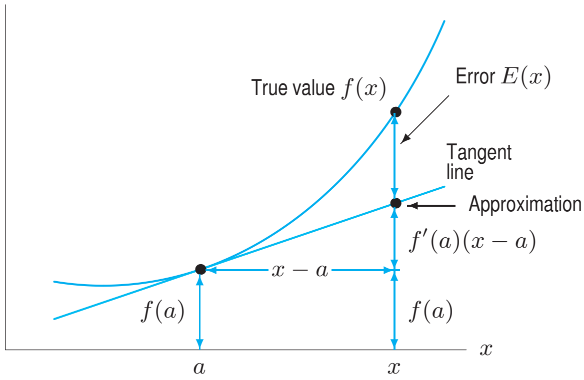

Figure E.16: The tangent line approximation and its error

Example 4 What is the tangent line approximation for $f(x) = \sin x$ near $x = 0$? Assume that $f'(0) = 1$.

Solution The tangent line approximation of f near $x = 0$ is

$$f(x) \approx f(0) + f'(0)(x-0).$$

If $f(x) = \sin x$, then $f(0) = \sin 0 = 0$. Using the given fact that $f'(0) = 1$, the approximation is

$$\sin x \approx x.$$

This means that, near $x = 0$, the function $f(x) = \sin x$ is well approximated by the function $y = x$. If we zoom in on the graphs of the functions $\sin x$ and x near the origin, we won't be able to tell them apart. (See Figure E.17.)

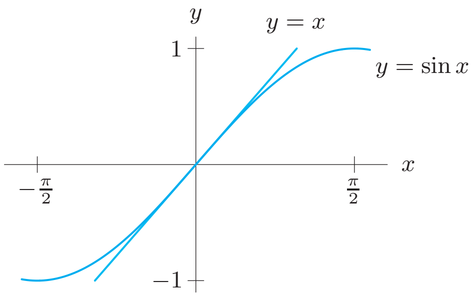

Figure E.17: Tangent line approximation to $y = \sin x$

Estimating the Error in the Approximation

Let us look at the error, $E(x)$, which is the difference between $f(x)$ and the local linearization. (Look back at Figure E.16.) The fact that the graph of f looks like a line as we zoom in means that not only is $E(x)$ small for x near a, but also that $E(x)$ is small relative to $(x - a)$. To demonstrate this, we prove the following theorem about the ratio $E(x)/(x - a)$.

Theorem: Differentiability and Local Linearity

Suppose f is differentiable at $x = a$ and $E(x)$ is the error in the tangent line approximation, that is:

$$E(x) = f(x) - f(a) - f'(a)(x - a).$$

Then

$$\lim_{x \to a} \frac{E(x)}{x - a} = 0.$$

Proof Using the definition of $E(x)$, we have

$$\frac{E(x)}{x - a} = \frac{f(x) - f(a) - f'(a)(x - a)}{x - a} = \frac{f(x) - f(a)}{x - a} - f'(a).$$

Taking the limit as $x \to a$ and using the definition of the derivative, we see that

$$\lim_{x \to a} \frac{E(x)}{x - a} = \lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a} - f'(a) \right) = f'(a) - f'(a) = 0.$$

Why Differentiability Makes A Graph Look Straight

We can use the error $E(x)$ to understand why differentiability makes a graph look straight when we zoom in.

Example 5 Consider the graph of $f(x) = \sin x$ near $x = 0$, and its linear approximation computed in Example 4. Show that there is an interval around 0 with the property that the distance from $f(x) = \sin x$ to the linear approximation is less than $0.1|x|$ for all x in the interval.

Solution The linear approximation of $f(x) = \sin x$ near 0 is $y = x$, so we write

$$\sin x = x + E(x).$$

Since $\sin x$ is differentiable at $x = 0$, the theorem tells us that

$$\lim_{x\to 0} \frac{E(x)}{x} = 0.$$

If we take $\epsilon = 1/10$, then the definition of limit guarantees that there is a $\delta > 0$ such that

$$\left|\frac{E(x)}{x}\right| < 0.1 \quad \text{for all} \quad |x| < \delta.$$

In other words, for x in the interval $(-\delta, \delta)$, we have $|x| < \delta$, so

$$|E(x)| < 0.1|x|.$$

(See Figure E.18.)

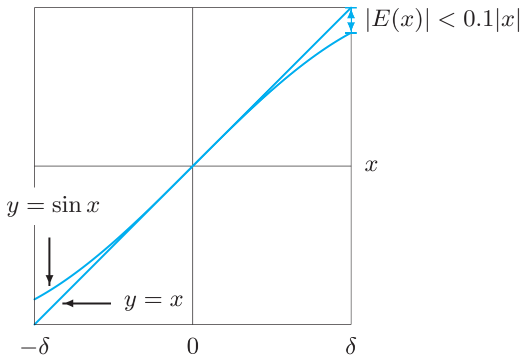

Figure E.18: Graph of $y = \sin x$ and its linear approximation $y = x$, showing a window in which the magnitude of the error, $|E(x)|$, is less than $0.1|x|$ for all x in the window

We can generalize from this example to explain why differentiability makes the graph of f look straight when viewed over a small graphing window. Suppose f is differentiable at $x = a$. Then we know $\lim\limits_{x\to a} \left|\dfrac{E(x)}{x-a}\right| = 0$. So, for any $\epsilon > 0$, we can find a δ small enough so that

$$\left|\frac{E(x)}{x-a}\right| < \epsilon, \qquad \text{for} \quad a-\delta < x < a+\delta.$$

So, for any x in the interval $(a-\delta, a+\delta)$, we have

$$|E(x)| < \epsilon|x-a|.$$

Thus, the error, $E(x)$, is less than ϵ times $|x-a|$, the distance between x and a. So, as we zoom in on the graph by choosing smaller ϵ, the deviation, $|E(x)|$, of f from its tangent line shrinks, even relative to the scale on the x-axis. So, zooming makes a differentiable function look straight.

Differentiability and Continuity

The fact that a function which is differentiable at a point has a tangent line suggests that the function is continuous there, as the next theorem shows.

> **Theorem: A Differentiable Function Is Continuous**
>
> If $f(x)$ is differentiable at a point $x = a$, then $f(x)$ is continuous at $x = a$.

Proof We assume $f(x)$ is differentiable at $x = a$. Then we know

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}.$$

So we must have

$$\lim_{x\to a}(f(x) - f(a)) = \lim_{x\to a}\left((x-a)\frac{f(x)-f(a)}{x-a}\right) = \left(\lim_{x\to a}(x-a)\right)\cdot\left(\lim_{x\to a}\frac{f(x)-f(a)}{x-a}\right)$$
$$= 0 \cdot f'(a) = 0.$$

Then,

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

which means $f(x)$ is continuous at $x = a$.

Problems for Section E

1. For each of the graphs in Figure E.19, list the x-values for which the function appears to be
(i) Not continuous and (ii) Not differentiable.

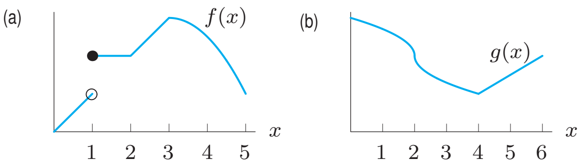

Figure E.19

2. Look at the graph of $f(x) = (x^2 + 0.0001)^{1/2}$ shown in Figure E.20. The graph of f appears to have a sharp corner at $x = 0$. Do you think f has a derivative at $x = 0$?

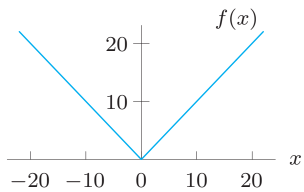

Figure E.20

Decide if the functions in Problems 3–5 are differentiable at $x = 0$. Try zooming in on a graphing calculator, or calculating the derivative $f'(0)$ from the definition.

3. $f(x) = (x + |x|)^2 + 1$

4. $f(x) = \begin{cases} x\sin(1/x) + x & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$

5. $f(x) = \begin{cases} x^2\sin(1/x) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases}$

6. An electric charge, Q, in a circuit is given as a function of time, t, by

$$Q = \begin{cases} C & \text{for } t \le 0 \\ Ce^{-t/RC} & \text{for } t > 0, \end{cases}$$

where C and R are positive constants. The electric current, I, is the rate of change of charge, so

$$I = \frac{dQ}{dt}.$$

(a) Is the charge, Q, a continuous function of time?
(b) Do you think the current, I, is defined for all times, t? [Hint: To graph this function, take, for example, $C = 1$ and $R = 1$.]

7. A magnetic field, B, is given as a function of the distance, r, from the center of a wire as follows:

$$B = \begin{cases} \dfrac{r}{r_0}B_0 & \text{for } r \le r_0 \\ \\ \dfrac{r_0}{r}B_0 & \text{for } r > r_0. \end{cases}$$

(a) Sketch a graph of B against r. What is the meaning of the constant B_0?
(b) Is B continuous at $r = r_0$? Give reasons.
(c) Is B differentiable at $r = r_0$? Give reasons.

8. A cable is made of an insulating material in the shape of a long, thin cylinder of radius r_0. It has electric charge distributed evenly throughout it. The electric field, E, at a distance r from the center of the cable is given by

$$E = \begin{cases} kr & \text{for} \quad r \le r_0 \\ k\dfrac{r_0^2}{r} & \text{for} \quad r > r_0. \end{cases}$$

(a) Is E continuous at r_0?
(b) Is E differentiable at r_0?
(c) Sketch a graph of E as a function of r.

9. Graph the function defined by

$$g(r) = \begin{cases} 1 + \cos(\pi r/2) & \text{for} \quad -2 \le r \le 2 \\ 0 & \text{for} \quad r < -2 \quad \text{or} \quad r > 2. \end{cases}$$

(a) Is g continuous at $r = 2$? Explain your answer.
(b) Do you think g is differentiable at $r = 2$? Explain your answer.

10. The potential, ϕ, of a charge distribution at a point on the y-axis is given by

$$\phi = \begin{cases} 2\pi\sigma\left(\sqrt{y^2 + a^2} - y\right) & \text{for } y \ge 0 \\ 2\pi\sigma\left(\sqrt{y^2 + a^2} + y\right) & \text{for } y < 0 \end{cases}$$

where σ and a are positive constants. [Hint: To graph this function, take, for example, $2\pi\sigma = 1$ and $a = 1$.]

(a) Is ϕ continuous at $y = 0$?
(b) Do you think ϕ is differentiable at $y = 0$?

11. Consider the function $f(x) = \sqrt{x}$. Assume $f'(4) = 1/4$.

(a) Find and sketch $f(x)$ and the tangent line approximation to $f(x)$ near $x = 4$.
(b) Compare the true value of $f(4.1)$ with the value obtained by using the tangent line approximation.
(c) Compare the true and approximate values of $f(16)$.
(d) Using a graph, explain why the tangent line approximation is a good one when $x = 4.1$ but not when $x = 16$.

12. Local linearization will give values too small for the function x^2 and too large for the function $\sqrt{x}$. Draw pictures to explain why.

13. Find the local linearization of $f(x) = x^2$ near $x = 1$.

14. Consider the graph of $f(x) = x^2$ near $x = 1$. Find an interval around $x = 1$ with the property that in any smaller interval, the graph of $f(x) = x^2$ never diverges from its local linearization by more than $0.1|x - 1|$ for all x in the interval.

15. Consider a function f and a point a. Suppose there is a number L such that the linear function g

$$g(x) = f(a) + L(x - a)$$

is a good approximation to f. By good approximation, we mean that

$$\lim_{x \to a} \frac{E_L(x)}{x - a} = 0,$$

where $E_L(x)$ is the approximation error defined by

$$f(x) = g(x) + E_L(x) = f(a) + L(x - a) + E_L(x).$$

Show that f is differentiable at $x = a$ and that $f'(a) = L$. Thus the tangent line approximation is the only good linear approximation.

F THE DEFINITE INTEGRAL

Recall that if f is continuous on $[a, b]$ the definite integral is given by a limit of left or right sums:

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_i)\Delta x = \lim_{x\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x.$$

This provides a method for approximating definite integrals numerically.[5] In this section we give a formal definition of the definite integral that makes use of more general sums.

A Special Case: Monotonic Functions

A function which is either increasing throughout an interval or decreasing throughout that interval is said to be *monotonic* on the interval. In Section 5.1 we saw that if f is monotonic, the left and right sums trap the exact value of the integral between them. Let us consider Example 1 on page 283 of the textbook, which looks at the value of

$$\int_1^2 \frac{1}{t}\,dt.$$

The left- and right-hand sums for $n = 2$, 10, 50, and 250 are listed in Table F.1.

Because the function $f(t) = 1/t$ is decreasing, the left-hand sums converge to the integral from above, and the right-hand sums converge from below. From the last row of the table we can deduce that

$$0.6921 < \int_1^2 \frac{1}{t}\,dt < 0.6941,$$

so $\int_1^2 \frac{1}{t}\,dt \approx 0.69$ to two decimal places.

The Difference Between the Upper and Lower Estimates

To be sure that the left- and right-hand sums trap a unique number between them, we need to know that the difference between them approaches zero. On page 277 we saw that for a monotonic function f on the interval $[a, b]$:

$$\left|\begin{array}{c}\text{Difference between}\\ \text{upper and lower estimates}\end{array}\right| = |f(b) - f(a)| \cdot \Delta t,$$

where $\Delta t = (b - a)/n$. We can make this difference as small as we like by choosing Δt small enough.

When f Is Not Monotonic

If f is not monotonic, the definite integral is not always bracketed between the left- and right-hand sums. For example, Table F.2 gives sums for the integral $\int_0^{2.5} \sin(t^2)\,dt$. Although $\sin(t^2)$ is certainly not monotonic on $[0, 2.5]$, by the time we get to $n = 250$, it is pretty clear that $\int_0^{2.5} \sin(t^2)\,dt \approx 0.43$ to two decimal places. Notice, however, that 0.43 does not lie between 1.2500 and 1.2085, the left- and right-hand sums for $n = 2$, or even between 0.4614 and 0.4531, the two sums for $n = 10$. If the integrand is not monotonic, the left- and right-hand sums may both be larger (or smaller) than the integral. (See Problems 2 and 3 for more examples of this behavior.)

Table F.1 *Left- and right-hand sums for* $\int_1^2 \frac{1}{t}\,dt$

n	Left-hand sum	Right-hand sum
2	0.8333	0.5833
10	0.7188	0.6688
50	0.6982	0.6882
250	0.6941	0.6921

[5] In practice, we often approximate integrals using more sophisticated numerical methods.

Table F.2 *Left- and right-hand sums for* $\int_0^{2.5} \sin(t^2)\, dt$

n	Left-hand sum	Right-hand sum
2	1.2500	1.2085
10	0.4614	0.4531
50	0.4324	0.4307
250	0.4307	0.4304
1000	0.4306	0.4305

Defining The Definite Integral by Upper and Lower Sums

When f is not monotonic, it is difficult to get upper and lower bounds for $\int_a^b f(x)\, dx$ from left and right sums. So instead we take the following approach for any function, f. If f is continuous, this new approach agrees with the previous approach. As before, we consider a subdivision of $[a, b]$ into n intervals; however, now we allow the subintervals to have different lengths. We let Δx_i be the length of the i-th interval, and make the following definition:

Suppose that f is bounded above and below on $[a, b]$. A **lower sum** for f on the interval $[a, b]$ is a sum

$$\sum_{i=1}^{n} m_i \Delta x_i,$$

where m_i is the greatest lower bound for f on the i-th interval. An **upper sum** is

$$\sum_{i=1}^{n} M_i \Delta x_i,$$

where M_i is the least upper bound for f on the i-th interval.

See Figure F.21. Now instead of taking a limit as $n \to \infty$, we consider the least upper and greatest lower bounds of these sums. We make the following definition.

Definition of the Definite Integral

Suppose that f is bounded above and below on $[a, b]$. Let L be the least upper bound for all the lower sums for f on $[a, b]$, and let U be the greatest lower bound for all the upper sums. If $L = U$, then we say that f is *integrable* and we define $\int_a^b f(x)\, dx$ to be equal to the common value of L and U.

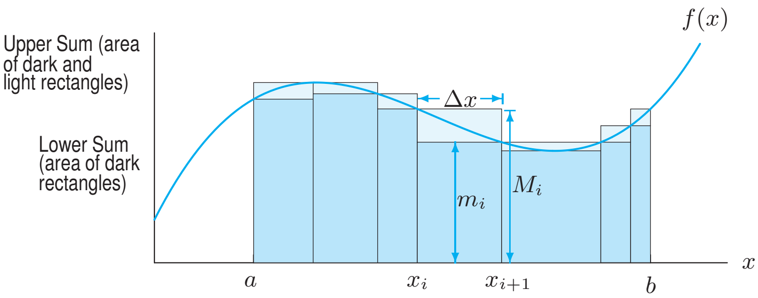

Figure F.21: Lower and upper sums approximating $\int_a^b f(x)\, dx$

Using the Definition in a Proof

As an example, we will prove the theorem stated on page 306 of the textbook:

Theorem: The Mean Value Inequality for Integrals

If $m \leq f(x) \leq M$ for all x in $[a, b]$, and if f is integrable on $[a, b]$, then

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$

Geometrically, if f is positive, this theorem says that the area under the graph of f is less than the area of the rectangle of height M, and greater than the area of the rectangle of height m. See Figure **??** on page **??** of the textbook.

Proof The simplest subdivision of $[a, b]$ is the one that consists of one subinterval, namely, $[a, b]$ itself. Although it does not give a very good approximation for the definite integral, it still counts as a subdivision. The least upper bound for f on $[a, b]$ is less than or equal to M, and the length of the only subinterval in the subdivision is $b-a$. So the upper sum for this subdivision is less than or equal to $M(b-a)$. Since every upper sum is an upper estimate for $\int_a^b f(x)\,dx$, we have

$$\int_a^b f(x)\,dx \leq \text{Upper sum} \leq M(b-a).$$

The argument for the other inequality is similar, using lower sums. (See Problem 19.)

Problem 20 gives another example of a proof using the definition of the definite integral.

Continuous Functions Are Integrable

In this section we will prove the following:

Theorem: Continuous Functions are Integrable

If f is continuous on $[a, b]$, then $\int_a^b f(x)\,dx$ exists.

The Key Question

It can be shown (for example, using the Extreme Value Theorem on page 196) that a continuous function on a finite interval is bounded above and below. Since any lower sum is less than or equal to any upper sum (see Problems 13–17), it follows that $L \leq U$. (See Problem 18.) To show that f is integrable, we need only show that $L = U$, so the question is the following:

Can we find a subdivision where the lower sum is as close as we like to the upper sum?

If we could, then $L < U$ would not be a possibility, since then $U - L$ would be a positive number, and we would be able to find lower and upper sums less than $U - L$ apart. In that case, either the lower sum would be bigger than L or the upper sum less than U. This can't happen, since L is an upper bound for the lower sums, and U is a lower bound for the upper sums.

We have already seen that if f is monotonic we can find upper and lower sums that are arbitrarily close; just take the left and right sums. This proves that monotonic functions are integrable. If f is not monotonic, we proceed differently.

Squeezing the Integral Between Lower and Upper Sums

For a subdivision of $[a, b]$ we have

$$\text{Difference between upper and lower sums} = \sum_{i=1}^{n}(M_i - m_i)\Delta x_i,$$

where M_i is the least upper bound for f on the i-th subinterval and m_i is the greatest lower bound. The number $M_i - m_i$ represents the amount by which f varies on the i-th subinterval; we call $M_i - m_i$ the variation[6] of f on this subinterval. Suppose that we could choose the subdivision so that the variation on each subinterval was less than some small positive number ϵ. Then

$$\text{Difference between upper and lower sums} < \sum_{i=1}^{n}\epsilon\Delta x_i = \epsilon\sum_{i=1}^{n}\Delta x_i = \epsilon(b-a).$$

By choosing ϵ small enough we would be able to make the difference as small as we liked. Thus the next question is:

> Can we find a subdivision where the maximum variation of f on each of the subintervals is as small as we like?

Making the Variation Small on Each Subinterval

Let ϵ be a positive number, as small as we like. We want to prove that there is a subdivision of $[a, b]$ such that the variation of f on each subinterval is less than ϵ. We will give an indirect proof: we assume that there is no such subdivision, and show that this leads to impossible consequences.

Divide the interval $[a, b]$ into two halves. If each half has a subdivision where the maximum variation on subintervals is less than ϵ, we can put the two subdivisions together to form a subdivision of $[a, b]$ with the same property.

So if $[a, b]$ fails to have such a subdivision, then so does one of the halves. Choose a half that does not have such a subdivision and divide it in half again. Once more, one of the halves must fail to have a subdivision where the variation of f on each subinterval is less than ϵ. Continuing in this way we find a nested sequence of intervals, each of which fails to have such a subdivision.

By the Nested Interval Theorem on page **??**, these intervals all contain some number c. Since f is continuous at c, we can find an interval around c on which the variation is less than ϵ. One of our nested intervals must be contained in this interval, since they get arbitrarily small. So the variation of f on one of the nested intervals is less than ϵ. This is impossible, given the way we chose each nested interval. So our supposition that $[a, b]$ fails to have the required subdivision is false; there must be a subdivision of $[a, b]$ such that the variation of f on each subinterval is less than ϵ.

Summary

We have shown by contradiction that for every positive number ϵ, no matter how small, there is a subdivision of $[a, b]$ such that the variation of f on each subinterval is less than ϵ. This means we can make the upper and lower sums as close as we like; hence $L = U$ and $\int_a^b f(x)\,dx = U = L$. That is, the continuous function f is integrable.

More General Riemann Sums

Left- and right-hand sums are special cases of Riemann sums. For a general Riemann sum, as with upper and lower sums, we allow subdivisions to have different lengths. Also, instead of evaluating

[6] This is not the same as the total variation used in more advanced texts.

f only at the left or right endpoint of each subdivision, we allow it to be evaluated anywhere in the subdivision. Thus, a general Riemann sum has the form

$$\sum_{i=1}^{n} (\text{Value of } f \text{ at some point in } i\text{-th subdivision}) \cdot (\text{Length of } i\text{-th subdivision}).$$

(See Figure F.22.) As before, we let $x_0, x_1, \ldots, x_n$ be the endpoints of the subdivisions, so the length of the i-th subdivision is $\Delta x_i = x_i - x_{i-1}$. For each i we choose a point c_i in the i-th subinterval at which to evaluate f, leading to the following definition:

A general Riemann sum for f on the interval $[a, b]$ is a sum of the form

$$\sum_{i=1}^{n} f(c_i)\Delta x_i,$$

where $a = x_0 < x_1 < \cdots < x_n = b$, and, for $i = 1, \ldots, n$, $\Delta x_i = x_i - x_{i-1}$, and $x_{i-1} \le c_i \le x_i$.

We define the *error* in an approximation to be the magnitude of the difference between the approximate and the true values. (Notice that error doesn't mean mistake here.) Since the true value of the integral is between any upper estimate and any lower estimate, the error in approximating a definite integral by a Riemann sum must be less than the difference between the upper and lower sums using the same subdivision. If $\int_a^b f(x)\,dx$ exists, there is a subdivision for which the upper and lower sums are as close as we like. So we can approximate the integral arbitrarily closely by Riemann sums.

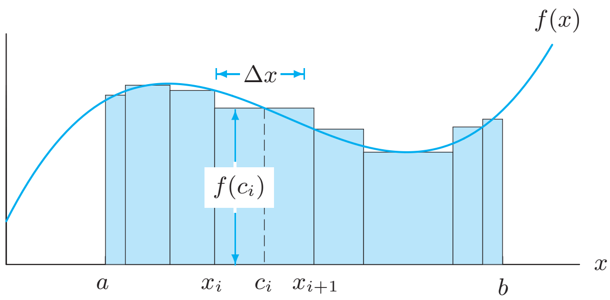

Figure F.22: A general Riemann sum approximating $\int_a^b f(x)\,dx$

Problems for Section F

1. Write a few sentences in support of or in opposition to the following statement:

"If a left-hand sum underestimates a definite integral by a certain amount, then the corresponding right-hand sum will overestimate the integral by the same amount."

2. Using the graph of $2 + \cos x$, for $0 \le x \le 4\pi$, list the following quantities in increasing order: the value of the integral $\int_0^{4\pi} (2+\cos x)\,dx$, the left-hand sum with $n = 2$ subdivisions, and the right-hand sum with $n = 2$ subdivisions.

3. Sketch the graph of a function f (you do not need to give a formula for f) on an interval $[a, b]$ with the property that with $n = 2$ subdivisions,

$$\int_a^b f(x)\,dx < \text{Left-hand sum } < \text{Right-hand sum}.$$

For Problems 4–12, find a subdivision using subintervals of equal length for which the lower and upper sums differ by less than 0.1. Give these sums and an estimate for the integral which is within 0.05 of the true value. Explain your reasoning. [Except for Problem 12, each function is monotonic over the given interval.]

4. $\displaystyle\int_0^5 x^2\,dx$

5. $\displaystyle\int_1^2 2^x\,dx$

6. $\displaystyle\int_1^4 \frac{1}{\sqrt{1+x^2}}\,dx$

7. $\displaystyle\int_1^{1.5} \sin x\,dx$

8. $\displaystyle\int_0^{\pi/4} \frac{d\theta}{\cos\theta}$

9. $\displaystyle\int_{-2}^{-1} \cos^3 y\,dy$

10. $\displaystyle\int_1^5 (\ln x)^2\,dx$

11. $\displaystyle\int_{1.1}^{1.7} e^t \ln t\,dt$

12. $\displaystyle\int_{-3}^{3} e^{-t^2}\,dt$

In Problems 13–17 you will show that every lower sum for a given bounded function f on an interval $[a, b]$ is less than every upper sum, using the idea of a *refinement* of a subdivision. Given a subdivision of the interval $[a, b]$, we can subdivide one or more of its subintervals to obtain a new subdivision. We say that the new subdivision is a *refinement* of the old one. Notice that if one subdivision's set of endpoints contains another's, then the first is a refinement of the second.

13. Show that the lower sum for f on $[a, b]$ using a given subdivision is less than or equal to the upper sum using the same subdivision.

14. In this problem we will show that refining a subdivision results in a lower sum which is larger than the original. Let f be a function defined and bounded from below on $[a, b]$, and choose a subdivision of $[a, b]$, with endpoints $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$.

(a) Suppose $x_{i-1} \le y \le x_i$. Let m_i be the greatest lower bound for f on $[x_{i-1}, x_i]$. Show that m_i is less than or equal to the greatest lower bound for f on $[x_{i-1}, y]$ and the greatest lower bound for f on $[y, x_i]$.

(b) Show that the lower sum for f using the subdivision $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ is less than or equal to the lower sum using the same subdivision with y included.

(c) Show that the lower sum for f using the subdivision $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ is less than or equal to the lower sum using any refinement of the subdivision.

15. In this problem we will show that refining a subdivision results in an upper sum which is smaller than the original. Let f be a function defined and bounded from above on $[a, b]$, and choose a subdivision of $[a, b]$, with endpoints $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$.

(a) Suppose $x_{i-1} \le y \le x_i$. Let M_i be the least upper bound for f on $[x_{i-1}, x_i]$. Show that M_i is greater than or equal to the least upper bound for f on $[x_{i-1}, y]$ and the least upper bound for f on $[y, x_i]$.

(b) Show that the upper sum for f using the subdivision $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ is greater than or equal to the upper sum using the same subdivision with y included.

(c) Show that the upper sum for f using the subdivision $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ is greater than or equal to the upper sum using any refinement of the subdivision.

16. Given two subdivisions of $[a, b]$, show that there is a third one which is a refinement of both.

17. Show that any lower sum for f on $[a, b]$ is less than or equal to any upper sum. [Hint: The lower sum uses one subdivision of $[a, b]$; the upper sum uses another. Use Problem 16 to choose a common refinement of the two subdivisions and then use Problems 13–15.]

18. Let f be a function defined and bounded on $[a, b]$, let L be the least upper bound for all the lower sums for f on $[a, b]$, and let U be the greatest lower bound for all the upper sums.

(a) Show that if L were strictly greater than U, then there would be a lower sum that was strictly greater than an upper sum. [Hint: Let $\epsilon = L - U$, and find a lower sum within $\epsilon/3$ of L and an upper sum within $\epsilon/3$ of U.]

(b) Deduce that $L \le U$.

19. On page **??** we proved one half of the Mean Value Inequality for Integrals. Prove the other half: that is, if f is continuous on $[a, b]$ and $f(x) \ge m$ for x in $[a, b]$, then $m(b-a) \le \int_a^b f(x)\,dx$.

20. In this problem we will prove that if f is continuous on $[a, b]$ and if c is in $[a, b]$, then

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

(a) Show that if ℓ_1 is a lower sum for f on $[a, c]$, and if ℓ_2 is a lower sum for f on $[c, b]$, then $\ell_1 + \ell_2$ is a lower sum for f on $[a, b]$.

(b) Show that if ℓ is a lower sum for f on $[a, b]$, then there is a lower sum ℓ_1 for f on $[a, c]$ and a lower sum ℓ_2 for f on $[c, b]$ such that $\ell \le \ell_1 + \ell_2$.

(c) Let L be the least upper bound of all the lower sums on $[a, b]$, let L_1 be the least upper bound of all the lower sums on $[a, c]$, and let L_2 be the least upper bound of all the lower sums on $[c, a]$. Use parts (a) and (b) to show that $L = L_1 + L_2$.

Since f is continuous on $[a, b]$, $L = \int_a^b f(x)\,dx$, $L_1 = \int_a^c f(x)\,dx$, and $L_2 = \int_c^b f(x)\,dx$. Thus you have proved the required statement.

G THEOREMS ABOUT CONTINUOUS AND DIFFERENTIABLE FUNCTIONS

In Chapter 4 of the textbook we used some basic facts without proof: for example, that a continuous function has a maximum on a bounded, closed interval, or that a function whose derivative is positive on an interval is increasing on that interval.

From a geometric point of view, these facts seem obvious. If we draw the graph of a continuous function, starting at one end of a bounded, closed interval and going to the other, it seems obvious that we must pass a highest point on the way. And if the derivative of a function is positive, then its graph must be sloping up, so the function has to be increasing.

However, this sort of graphical reasoning is not a rigorous proof, for two reasons. First, no matter how many pictures we imagine, we can't be sure we have covered all possibilities. Second, our pictures often depend on the theorems we are trying to prove.

A Continuous Function on a Closed Interval Has a Maximum

The Extreme Value Theorem
If f is continuous on the interval $[a, b]$, then f has a global maximum and a global minimum on that interval.

Our proof has two parts: The first is to show that f has an upper bound on $[a, b]$, the second is to show that if f has an upper bound then it has a global maximum on the interval. Here we prove the second part; the first part is proved in Problems 15 and 16. Then in Problem 5 we extend the result from maxima to minima.

Proof We assume that f is continuous and has an upper bound on the interval $[a, b]$. This means f has a least upper bound M on $[a, b]$. Divide $[a, b]$ into two halves. Then, on one of the halves, the least upper bound for f is M, for if it were less than M on both halves, it would be less than M on the whole. Choose a half on which the least upper bound is equal to M. Continue bisecting and at each stage choose the half-interval where the least upper bound for f is M. See Figure G.23. This results in a sequence of nested intervals. By the Nested Interval Theorem on page **??**, there is a number c in $[a, b]$ which is contained in all these intervals.

Since M is the least upper bound for f, we have $f(c) \leq M$. It is not possible that $f(c) < M$. For if $f(c) < M$, then $f(c) < M_0$ for some number $M_0 < M$. (For example, we could take M_0 to be half-way between M and $f(c)$.) But then, since f is continuous, there would be a $\delta > 0$ such that $f(x) < M_0$ for all x in $[a, b]$ with $c - \delta < x < c + \delta$. (See Problem 14.) Since the nested intervals we constructed above have width tending to zero, one of them would be contained in the interval $c - \delta < x < c + \delta$. Therefore, f would be bounded above by M_0 on one of the nested

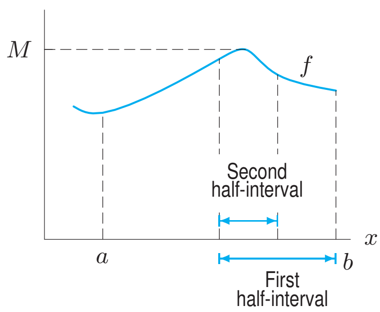

Figure G.23: Successively choosing the half-interval where the least upper bound of f is M

intervals. However, we chose each nested interval so that the least upper bound for f is M. This is a contradiction of $M_0 < M$.

So it is not possible that $f(c) < M$; we must have $f(c) = M$. Thus, M is the global maximum of f on $[a, b]$, which is what we wanted to show.

The Extreme Value Theorem guarantees the existence of global maxima (and minima) on an interval. To actually find the global maxima, we look at all the local maxima. The following theorem tells us that inside an interval, local maxima only occur at critical points, where the derivative is either zero or undefined.

Theorem: Local Extrema and Critical Points
Suppose f is defined on an interval and has a local maximum or minimum at the point $x = a$, which is not an endpoint of the interval. If f is differentiable at $x = a$, then $f'(a) = 0$.

Proof We start with the definition of the derivative:

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}.$$

Remember that this is a two-sided limit:

$$f'(a) = \lim_{h\to 0^-} \frac{f(a+h) - f(a)}{h} = \lim_{h\to 0^+} \frac{f(a+h) - f(a)}{h}.$$

Suppose that f has a local maximum at $x = a$. By the definition of local maximum, $f(a+h) \le f(a)$ for all sufficiently small h. Thus $f(a+h) - f(a) \le 0$ for sufficiently small h. The denominator, h, is positive when we take the limit from the right and negative when we take the limit from the left. Thus

$$\lim_{h\to 0^-} \frac{f(a+h) - f(a)}{h} \ge 0 \quad \text{and} \quad \lim_{h\to 0^+} \frac{f(a+h) - f(a)}{h} \le 0.$$

Since both these limits are equal to $f'(a)$, we have $f'(a) \ge 0$ and $f'(a) \le 0$, so we must have $f'(a) = 0$.

A Relationship Between Local and Global: The Mean Value Theorem

We often want to infer a global conclusion (for example, f is increasing on an interval) from local information (f' is positive.) The following theorem relates the average rate of change of a function on an interval (global information) to the instantaneous rate of change at a point in the interval (local information).

The Mean Value Theorem
If f is continuous on $[a, b]$ and differentiable on (a, b), then there exists a number c, with $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

In other words, $f(b) - f(a) = f'(c)(b - a)$.

To understand what this theorem is saying geometrically, consider the graph in Figure G.24. Join the points on the curve where $x = a$ and $x = b$ with a line and observe that the slope of this secant line AB is given by

$$m = \frac{f(b) - f(a)}{b - a}.$$

Now consider the tangent line drawn to the curve at each point between $x = a$ and $x = b$. In general, these lines will have different slopes. For the curve shown in Figure G.24, the tangent line at $x = a$ is flatter than the secant line from A to B. Similarly, the tangent line at $x = b$ is steeper than the secant line. However, there is at least one point between a and b where the slope of the tangent line to the curve is precisely the same as the slope of the secant line. Suppose this occurs at $x = c$. Then

$$f'(c) = m = \frac{f(b) - f(a)}{b - a}.$$

The Mean Value Theorem tells us that the point $x = c$ exists, but it does not tell us how to find c.

Problems 17 and 18 show how the Mean Value Theorem can be deduced from the Extreme Value Theorem.

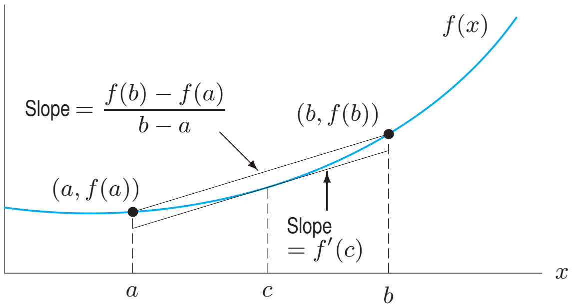

Figure G.24: The point c with $f'(c) = \frac{f(b)-f(a)}{b-a}$

The Increasing Function Theorem

We say that a function f is *increasing* on an interval if, for any two numbers x_1 and x_2 in the interval such that $x_1 < x_2$, we have $f(x_1) < f(x_2)$. If instead we have $f(x_1) \leq f(x_2)$, we say f is *nondecreasing*.

The Increasing Function Theorem

Suppose that f is continuous on $[a, b]$ and differentiable on (a, b).

- If $f'(x) > 0$ on (a, b), then f is increasing on $[a, b]$.
- If $f'(x) \geq 0$ on (a, b), then f is nondecreasing on $[a, b]$.

Proof Suppose $a \leq x_1 < x_2 \leq b$. By the Mean Value Theorem, there is a number c, with $x_1 < c < x_2$, such that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1).$$

If $f'(c) > 0$, this says $f(x_2) - f(x_1) > 0$, which means f is increasing. If $f'(c) \geq 0$, this says $f(x_2) - f(x_1) \geq 0$, which means f is nondecreasing.

It may seem that something as simple as the Increasing Function Theorem should follow immediately from the definition of the derivative, and that the use of the Mean Value Theorem (which in turn depends on the Extreme Value Theorem) is surprising. It is possible to give a proof which does not use the Mean Value Theorem, but not a simple one.

The Constant Function Theorem

If f is constant on an interval, then we know that $f'(x) = 0$ on the interval. The following theorem is the converse.

The Constant Function Theorem
Suppose that f is continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) = 0$ on (a, b), then f is constant on $[a, b]$.

Proof The proof is the same as for the Increasing Function Theorem, only in this case $f'(c) = 0$ so $f(x_2) - f(x_1) = 0$. Thus $f(x_2) = f(x_1)$ for $a \le x_1 < x_2 \le b$, so f is constant.

A proof of the Constant Function Theorem using the Increasing Function Theorem is given in Problems 6 and 8.

The Racetrack Principle

The Racetrack Principle[7]
Suppose that g and h are continuous on $[a, b]$ and differentiable on (a, b), and that $g'(x) \le h'(x)$ for $a < x < b$.

- If $g(a) = h(a)$, then $g(x) \le h(x)$ for $a \le x \le b$.
- If $g(b) = h(b)$, then $g(x) \ge h(x)$ for $a \le x \le b$.

The Racetrack Principle has the following interpretation. We can think of $g(x)$ and $h(x)$ as the positions of two racehorses at time x, with horse h always moving faster than horse g. If they start together, horse h is ahead during the whole race. If they finish together, horse g was ahead during the whole race.

Proof Consider the function $f(x) = h(x) - g(x)$. Since $f'(x) = h'(x) - g'(x) \ge 0$, we know that f is nondecreasing by the Increasing Function Theorem. So $f(x) \ge f(a) = h(a) - g(a) = 0$. Thus $g(x) \le h(x)$ for $a \le x \le b$. This proves the first part of the Racetrack Principle. Problem 7 asks for a proof of the second part.

Example 1 Explain graphically why $e^x \ge 1 + x$ for all values of x. Then use the Racetrack Principle to prove the inequality.

Solution The graph of the function $f(x) = e^x$ is concave up everywhere and the equation of its tangent line at the point $(0, 1)$ is $y = x + 1$. (See Figure G.25.) Since the graph always lies above its tangent, we have the inequality

$$e^x \ge 1 + x.$$

Now we prove the inequality using the Racetrack Principle. Let $g(x) = 1 + x$ and $h(x) = e^x$. Then $g(0) = h(0) = 1$. Furthermore, $g'(x) = 1$ and $h'(x) = e^x$. Hence $g'(x) \le h'(x)$ for $x \ge 0$. So by the Racetrack Principle, with $a = 0$, we have $g(x) \le h(x)$, that is, $1 + x \le e^x$.

For $x \le 0$ we have $h'(x) \le g'(x)$. So by the Racetrack Principle, with $b = 0$, we have $g(x) \le h(x)$, that is, $1 + x \le e^x$.

[7]Based on the Racetrack Principle in *Calculus&Mathematica*, by William Davis, Horacio Porta, Jerry Uhl (Reading: Addison Wesley, 1994).

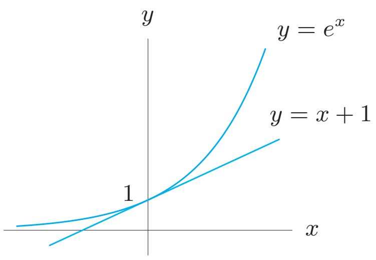

Figure G.25: Graph showing that $e^x \geq 1 + x$

Problems for Section G

1. Use the Racetrack Principle and the fact that $\sin 0 = 0$ to show that $\sin x \leq x$ for all $x \geq 0$.

2. Use the Racetrack Principle to show that $\ln x \leq x - 1$.

3. Use the fact that $\ln x$ and e^x are inverse functions to show that the inequalities $e^x \geq 1 + x$ and $\ln x \leq x - 1$ are equivalent for $x > 0$.

4. Suppose that the position of a particle moving along the x-axis is given by $s = f(t)$, and that the initial position and velocity of the particle are $f(0) = 3$ and $f'(0) = 4$. Suppose that the acceleration is bounded by $5 \leq f''(t) \leq 7$ for $0 \leq t \leq 2$. What can we say about the position $f(2)$ of the particle at $t = 2$?

5. Show that if every continuous function on an interval $[a, b]$ has a global maximum, then every continuous function has a global minimum as well. [Hint: Consider $-f$.]

6. State a Decreasing Function Theorem, analogous to the Increasing Function Theorem. Deduce your theorem from the Increasing Function Theorem. [Hint: Apply the Increasing Function Theorem to $-f$.]

7. Suppose that g and h are continuous on $[a, b]$ and differentiable on (a, b). Prove that if $g'(x) \leq h'(x)$ for $a < x < b$ and $g(b) = h(b)$, then $h(x) \leq g(x)$ for $a \leq x \leq b$.

8. Deduce the Constant Function Theorem from the Increasing Function Theorem and the Decreasing Function Theorem (see problem 6).

9. Prove that if $f'(x) = g'(x)$ for all x in (a, b), then there is a constant C such that $f(x) = g(x) + C$ on (a, b). [Hint: Apply the Constant Function Theorem to $h(x) = f(x) - g(x)$.]

10. Suppose that $f'(x) = f(x)$ for all x. Prove that $f(x) = Ce^x$ for some constant C. [Hint: Consider the function $f(x)/e^x$.]

11. Suppose that f is continuous on $[a, b]$ and differentiable on (a, b) and that $m \leq f'(x) \leq M$ on (a, b). Use the Racetrack Principle to prove that $f(x) - f(a) \leq M(x - a)$ for all x in $[a, b]$, and that $m(x - a) \leq f(x) - f(a)$ for all x in $[a, b]$. Conclude that $m \leq (f(b) - f(a))/(b - a) \leq M$. This is called the Mean Value Inequality. In words: If the instantaneous rate of change of f is between m and M on an interval, so is the average rate of change of f over the interval.

12. Suppose that $f''(x) \geq 0$ for all x in (a, b). We will show the graph of f lies above the tangent line at $(c, f(c))$ for any c with $a < c < b$.

(a) Use the Increasing Function Theorem to prove that $f'(c) \leq f'(x)$ for $c \leq x < b$ and that $f'(x) \leq f'(c)$ for $a < x \leq c$.

(b) Use (a) and the Racetrack Principle to conclude that $f(c) + f'(c)(x - c) \leq f(x)$, for $a < x < b$.

13. In this problem we use the Mean Value Theorem to give a proof of the Fundamental Theorem of Calculus. Let f be continuous, with antiderivative F.

(a) Let $[a, b]$ be an interval contained in the domain of f, and let

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

be a subdivision of $[a, b]$. Show that there is a Riemann sum for f using this subdivision which is equal to $F(b) - F(a)$. [Hint: Apply the Mean Value Theorem to

$$\begin{aligned} F(b) - F(a) = ((&F(b) - F(x_{n-1})) \\ &+(F(x_{n-1}) - F(x_{n-2})) + \cdots \\ &+(F(x_1) - F(a)). \end{aligned}$$

(b) Deduce that $F(b) - F(a) = \int_a^b f(x)\, dx$.

14. Suppose that f is continuous on $[a, b]$, and let c be in $[a, b]$. Show that if $f(c) < M$, then there is a δ such that $f(x) < M$ for all x in $[a, b]$ such that $c-\delta < x < c+\delta$. [Hint: Let $\epsilon = M - f(c)$, and choose δ such that $|f(x) - f(c)| < \epsilon$ if $|x - c| < \delta$.]

On page 196 we proved that a continuous function f has a global maximum on the interval $[a, b]$ under the assumption that f has an upper bound on $[a, b]$. In Problems 15–16 we prove this claim.

15. **(a)** Suppose that f has no upper bound on $[a, b]$. Bisect $[a, b]$ into two halves. Deduce that f has no upper bound on at least one of the halves. Call that half $[a_1, b_1]$.

(b) Continue bisecting so that at the n^{th} stage you obtain an interval $[a_n, b_n]$ on which f has no upper bound. By the Nested Interval Theorem on page **??**, there is a point c in all the intervals $[a_n, b_n]$.

(c) Use continuity of f at c to deduce that f has an upper bound on $[a_n, b_n]$ for n sufficiently large. This contradicts the original supposition, so f must have an upper bound on $[a, b]$.

16. **(a)** Show that if $y \geq 0$, then $y/(1 + y) < 1$.

(b) Suppose that f is continuous on $[a, b]$ and that $f(x) \geq 0$ on $[a, b]$. Define a function g by $g(x) = f(x)/(1 + f(x))$. Show that g is continuous and bounded on $[a, b]$. It follows from the partial proof of the Extreme Value Theorem on page 196 that g has a global maximum on $[a, b]$ at some point $x = c$.

(c) Suppose that $y_1 \geq 0$ and $y_2 \geq 0$, and that $y_1/(1 + y_1) \leq y_2/(1 + y_2)$. Show that $y_1 \leq y_2$.

(d) Use parts (c) and (d) to show that f has a global maximum at $x = c$.

(e) We have shown that if f is continuous and non-negative on $[a, b]$, then it is bounded above on $[a, b]$. Now suppose that f is continuous, but not necessarily non-negative. By applying the argument to $|f|$, deduce that f is also bounded above.

17. In this problem we prove a special case of the Mean Value Theorem where $f(a) = f(b) = 0$. This special case is called Rolle's Theorem: If f is continuous on $[a, b]$ and differentiable on (a, b), and if $f(a) = f(b) = 0$, then there is a number c, with $a < c < b$, such that

$$f'(c) = 0.$$

By the Extreme Value Theorem, f has a global maximum and a global minimum on $[a, b]$.

(a) Prove Rolle's theorem in the case that both the global maximum and the global minimum are at endpoints of $[a, b]$. [Hint: $f(x)$ must be a very simple function in this case.]

(b) Prove Rolle's theorem in the case that either the global maximum or the global minimum is not at an endpoint. [Hint: Think about local maxima and minima.]

18. Use Rolle's Theorem to prove the Mean Value Theorem. Suppose that $f(x)$ is continuous on $[a, b]$ and differentiable on (a, b).

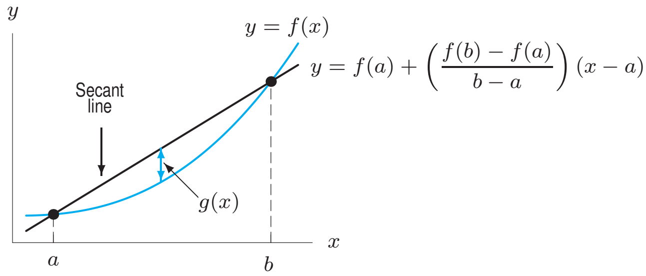

Figure G.26: $g(x)$ is the difference between the secant line and the graph of $f(x)$

(a) Let $g(x)$ be the difference between $f(x)$ and the y-value on the secant line joining $(a, f(a))$ to $(b, f(b))$. See Figure G.26. Show that

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

(b) Use Rolle's Theorem to show that there must be a point c in (a, b) such that $g'(c) = 0$.

(c) Show that if c is the point in part (b), then

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

H LIMITS AND CONTINUITY FOR FUNCTIONS OF MANY VARIABLES

The sheer vertical face of Half Dome, in Yosemite National Park in California, was caused by glacial activity during the Ice Age. (See Figure H.27.) The height of the terrain rises abruptly by nearly 1000 feet as we scale the rock from the west, whereas it is possible to make a gradual climb to the top from the east.

If we consider the function h giving the height of the terrain above sea level in terms of longitude and latitude, then h has a *discontinuity* along the path at the base of the cliff of Half Dome. Looking at the contour map of the region in Figure H.28, we see that in most places a small change in position results in a small change in height, except near the cliff. There, no matter how small a

Figure H.27: Half Dome in Yosemite National Park

Figure H.28: A contour map of Half Dome

step we take, we get a large change in height. (You can see how crowded the contours get near the cliff; some end abruptly along the discontinuity.)

This geological feature illustrates the ideas of continuity and discontinuity. Roughly speaking, a function is said to be *continuous* at a point if its values at places near the point are close to the value at the point. If this is not the case, the function is said to be *discontinuous*.

The property of continuity is one that, practically speaking, we usually assume of the functions we are studying. Informally, we expect (except under special circumstances) that values of a function do not change drastically when making small changes to the input variables. Whenever we model a one-variable function by an unbroken curve, we are making this assumption. Even when functions come to us as tables of data, we usually make the assumption that the missing function values between data points are close to the measured ones.

In this section we study limits and continuity a bit more formally in the context of functions of several variables. For simplicity we study these concepts for functions of two variables, but our discussion can be adapted to functions of three or more variables.

One can show that sums, products, and compositions of continuous functions are continuous, while the quotient of two continuous functions is continuous everywhere the denominator function is nonzero. Thus, each of the functions

$$\cos(x^2y), \qquad \ln(x^2+y^2), \qquad \frac{e^{x+y}}{x+y}, \qquad \ln(\sin(x^2+y^2))$$

is continuous at all points (x, y) where it is defined. As for functions of one variable, the graph of a continuous function over an unbroken domain is unbroken—that is, the surface has no holes or rips in it.

Example 1 From Figures H.29–H.32, which of the following functions appear to be continuous at $(0, 0)$?

(a) $f(x,y) = \begin{cases} \dfrac{x^2y}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$ (b) $g(x,y) = \begin{cases} \dfrac{x^2}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$

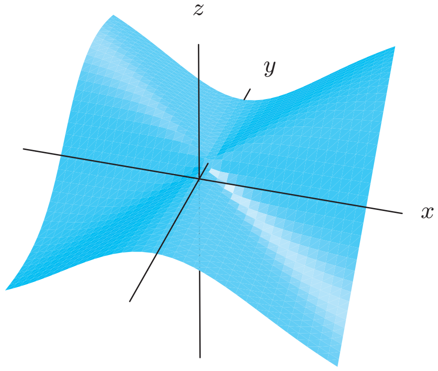

Figure H.29: Graph of $z = x^2y/(x^2+y^2)$

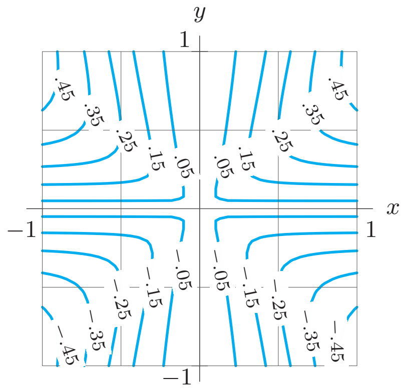

Figure H.30: Contour diagram of $z = x^2y/(x^2+y^2)$

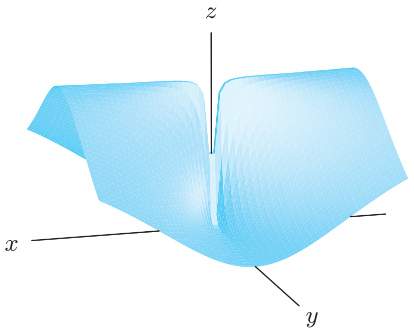

Figure H.31: Graph of $z = x^2/(x^2+y^2)$

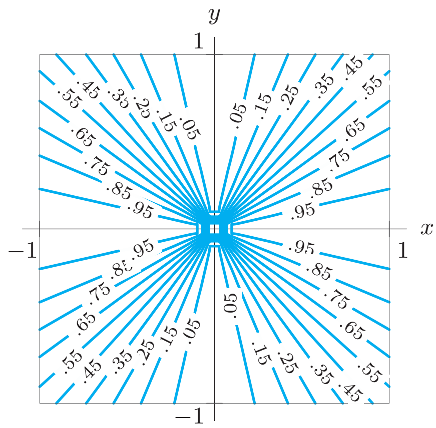

Figure H.32: Contour diagram of $z = x^2/(x^2+y^2)$

Solution

(a) The graph and contour diagram of f in Figures H.29 and H.30 suggest that f is close to 0 when (x, y) is close to $(0, 0)$. That is, the figures suggest that f is continuous at the point $(0, 0)$; the graph appears to have no rips or holes there.

However, the figures cannot tell us for sure whether f is continuous. To be certain we must investigate the limit analytically, as is done in Example 2(a) on page 40.

(b) The graph of g and its contours near $(0, 0)$ in Figure H.31 and H.32 suggest that g behaves differently from f: The contours of g seem to "crash" at the origin and the graph rises rapidly from 0 to 1 near $(0, 0)$. Small changes in (x, y) near $(0, 0)$ can yield large changes in g, so we expect that g is not continuous at the point $(0, 0)$. Again, a more precise analysis is given in Example 2(b) on page 40.

The previous example suggests that continuity *at* a point depends on a function's behavior *near* the point. To study behavior near a point more formally we need to define the limit of a function of two variables. Suppose that $f(x, y)$ is a function defined on a set in 2-space, not necessarily containing the point (a, b), but containing points (x, y) arbitrarily close to (a, b); suppose that L is a number.

The function f has a **limit** L at the point (a, b), written

$$\lim_{(x,y)\to(a,b)} f(x, y) = L,$$

if the difference $|f(x, y) - L|$ is as small as we wish whenever the distance from the point (x, y) to the point (a, b) is sufficiently small, but not zero.

We define continuity for functions of two variables in the same way as for functions of one variable:

A function f is **continuous at the point** (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

A function is **continuous** if it is continuous at each point of its domain.

Thus, if f is continuous at the point (a, b), then f must be defined at (a, b) and the limit, $\lim_{(x,y)\to(a,b)} f(x, y)$, must exist and be equal to the value $f(a, b)$. If a function is defined at a point (a, b) but is not continuous there, then we say that f is *discontinuous* at (a, b).

We now apply the definition of continuity to the functions in Example 1, showing that f is continuous at $(0, 0)$ and that g is discontinuous at $(0, 0)$.

Example 2 Let f and g be the functions defined everywhere on 2-space except at the origin as follows (a) $f(x, y) = \dfrac{x^2y}{x^2 + y^2}$ (b) $g(x, y) = \dfrac{x^2}{x^2 + y^2}$

Use the definition of the limit to show that $\lim_{(x,y)\to(0,0)} f(x, y) = 0$ and that $\lim_{(x,y)\to(0,0)} g(x, y)$ does not exist.

Solution (a) The graph and contour diagram of f both suggest that $\lim_{(x,y)\to(0,0)} f(x, y) = 0$. To use the definition of the limit, we must estimate $|f(x, y) - L|$ with $L = 0$:

$$|f(x, y) - L| = \left|\frac{x^2y}{x^2 + y^2} - 0\right| = \left|\frac{x^2}{x^2 + y^2}\right| |y| \le |y| \le \sqrt{x^2 + y^2},$$

Now $\sqrt{x^2 + y^2}$ is the distance from (x, y) to $(0, 0)$. Thus, to make $|f(x, y) - 0| < 0.001$, for example, we need only require (x, y) be within 0.001 of $(0, 0)$. More generally, for any positive number u, no matter how small, we are sure that $|f(x, y) - 0| < u$ whenever (x, y) is no farther than u from $(0, 0)$. This is what we mean by saying that the difference $|f(x, y) - 0|$ can be made as small as we wish by choosing the distance to be sufficiently small. Thus, we conclude that

$$\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^2 + y^2} = 0.$$

Notice that the function f has a limit at the point $(0, 0)$ even though f was not defined at $(0, 0)$. To make f continuous at $(0, 0)$ we must define its value there to be 0, as we did in Example 1.

(b) Although the formula defining the function g looks similar to that of f, we saw in Example 1 that g's behavior near the origin is quite different. If we consider points $(x, 0)$ lying along the x-axis near $(0, 0)$, then the values $g(x, 0)$ are equal to 1, while if we consider points $(0, y)$ lying along the y-axis near $(0, 0)$, then the values $g(0, y)$ are equal to 0. Thus, within any disk (no matter how small) centered at the origin, there are points where $g = 0$ and points where $g = 1$. Therefore the limit $\lim_{(x,y)\to(0,0)} g(x, y)$ does not exist.

While the notions of limit and continuity look formally the same for one- and two-variable functions, they are somewhat more subtle in the multivariable case. The reason for this is that on the line (1-space), we can approach a point from just two directions (left or right) but in 2-space there are an infinite number of ways to approach a given point.

Problems for Section H

1. Show that the function f does not have a limit at $(0,0)$ by examining the limits of f as $(x,y) \to (0,0)$ along the curve $y = kx^2$ for different values of k. The function is given by

$$f(x,y) = \frac{x^2}{x^2+y}, \qquad x^2 + y \neq 0.$$

2. Show that the function f does not have a limit at $(0,0)$ by examining the limits of f as $(x,y) \to (0,0)$ along the line $y = x$ and along the parabola $y = x^2$. The function is given by

$$f(x,y) = \frac{x^2 y}{x^4+y^2}, \qquad (x,y) \neq (0,0).$$

3. Consider the following function:

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

(a) Use a computer to draw the graph and the contour diagram of f.

(b) Do your answers to part (a) suggest that f is continuous at $(0,0)$? Explain your answer.

4. Consider the function f, whose graph and contour diagram are in Figures H.33 and H.34, and which is given by

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

(a) Show that $f(0,y)$ and $f(x,0)$ are each continuous functions of one variable.

(b) Show that rays emanating from the origin are contained in contours of f.

(c) Is f continuous at $(0,0)$?

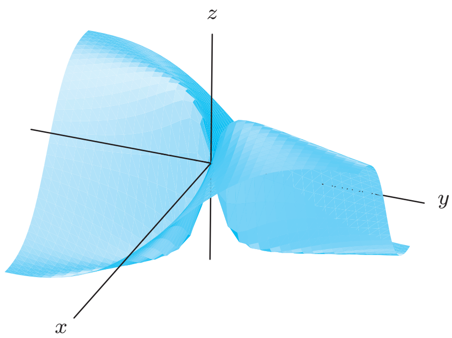

Figure H.33: Graph of $z = xy/(x^2+y^2)$

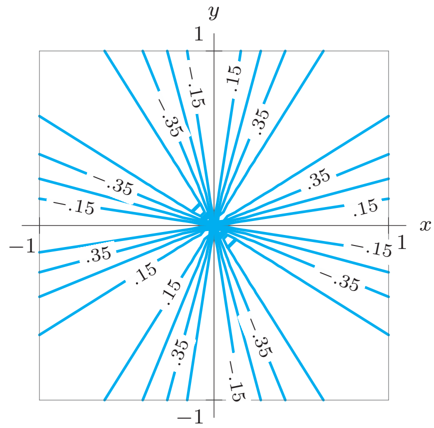

Figure H.34: Contour diagram of $z = xy/(x^2+y^2)$

For Problems 5–9 compute the limits of the functions $f(x,y)$ as $(x,y) \to (0,0)$. You may assume that polynomials, exponentials, logarithmic, and trigonometric functions are continuous.

5. $f(x,y) = x^2 + y^2$

6. $f(x,y) = e^{-x-y}$

7. $f(x,y) = \dfrac{x}{x^2+1}$

8. $f(x,y) = \dfrac{x+y}{(\sin y)+2}$

9. $f(x,y) = \dfrac{\sin(x^2+y^2)}{x^2+y^2}$ [Hint: You may assume that $\lim_{t\to 0}(\sin t)/t = 1$.]

For the functions in Problems 10–12, show that $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

10. $f(x,y) = \dfrac{x+y}{x-y}, \qquad x \neq y$

11. $f(x,y) = \dfrac{x^2-y^2}{x^2+y^2}$

12. $f(x,y) = \dfrac{xy}{|xy|}, \qquad x \neq 0$ and $y \neq 0$

13. Show that the contours of the function g defined in Example 1(b) on page 38 are rays emanating from the origin. Find the slope of the contour $g(x,y)=c$.

14. Explain why the following function is not continuous along the line $y=0$.

$$f(x,y) = \begin{cases} 1-x, & y \geq 0, \\ -2, & y < 0, \end{cases}$$

In Problems 15–16, determine whether there is a value for c making the function continuous everywhere. If so, find it. If not, explain why not.

15. $f(x,y) = \begin{cases} c+y, & x \leq 3, \\ 5-y, & x > 3. \end{cases}$

16. $f(x,y) = \begin{cases} c+y, & x \leq 3, \\ 5-x, & x > 3. \end{cases}$

I DIFFERENTIABILITY FOR FUNCTIONS OF MANY VARIABLES

In Section 14.3 of the textbook we gave an informal introduction to the concept of differentiability. We called a function $f(x,y)$ *differentiable* at a point (a,b) if it is well-approximated by a linear function near (a,b). This section focuses on the precise meaning of the phrase "well-approximated." By looking at examples, we shall see that local linearity requires the existence of partial derivatives, but they do not tell the whole story. In particular, existence of partial derivatives at a point is not sufficient to guarantee local linearity at that point.

We begin by discussing the relation between continuity and differentiability. As an illustration, take a sheet of paper, crumple it into a ball and smooth it out again. Wherever there is a crease it would be difficult to approximate the surface by a plane—these are points of nondifferentiability of the function giving the height of the paper above the floor. Yet the sheet of paper models a graph which is continuous—there are no breaks. As in the case of one-variable calculus, continuity does not imply differentiability. But differentiability does *require* continuity: there cannot be linear approximations to a surface at points where there are abrupt changes in height.

Differentiability For Functions Of Two Variables

For a function of two variables, as for a function of one variable, we define differentiability at a point in terms of the error and the distance from the point. If the point is (a,b) and a nearby point is $(a+h, b+k)$, the distance between them is $\sqrt{h^2+k^2}$. (See Figure I.35.)

A function $f(x,y)$ is **differentiable at the point** (a,b) if there is a linear function $L(x,y) = f(a,b) + m(x-a) + n(y-b)$ such that if the *error* $E(x,y)$ is defined by

$$f(x,y) = L(x,y) + E(x,y),$$

and if $h = x-a$, $k = y-b$, then the *relative error* $E(a+h,b+k)/\sqrt{h^2+k^2}$ satisfies

$$\lim_{\substack{h\to 0\\ k\to 0}} \frac{E(a+h,b+k)}{\sqrt{h^2+k^2}} = 0.$$

The function f is **differentiable** if it is differentiable at each point of its domain. The function $L(x,y)$ is called the *local linearization* of $f(x,y)$ near (a,b).

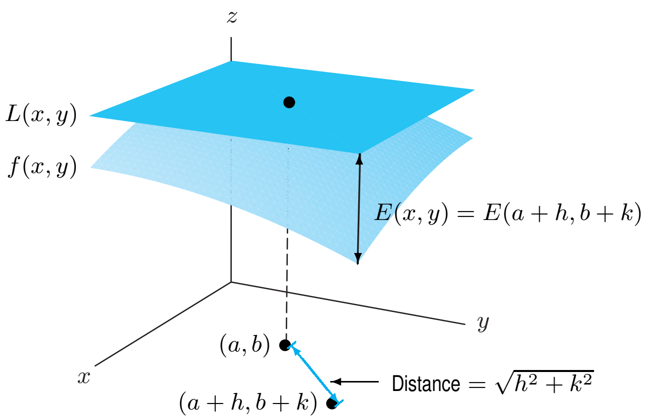

Figure I.35: Graph of function $z = f(x, y)$ and its local linearization $z = L(x, y)$ near the point (a, b)

Partial Derivatives and Differentiability

In the next example, we show that this definition of differentiability is consistent with our previous notion — that is, that $m = f_x$ and $n = f_y$ and that the graph of $L(x, y)$ is the tangent plane.

Example 1 Show that if f is a differentiable function with local linearization $L(x, y) = f(a, b) + m(x - a) + n(y - b)$, then $m = f_x(a, b)$ and $n = f_y(a, b)$.

Solution Since f is differentiable, we know that the relative error in $L(x, y)$ tends to 0 as we get close to (a, b). Suppose $h > 0$ and $k = 0$. Then we know that

$$\begin{aligned}
0 = \lim_{h\to 0} \frac{E(a+h,b)}{\sqrt{h^2+k^2}} = \lim_{h\to 0} \frac{E(a+h,b)}{h} &= \lim_{h\to 0} \frac{f(a+h,b) - L(a+h,b)}{h} \\
&= \lim_{h\to 0} \frac{f(a+h,b) - f(a,b) - mh}{h} \\
&= \lim_{h\to 0} \left(\frac{f(a+h,b) - f(a,b)}{h} \right) - m = f_x(a,b) - m.
\end{aligned}$$

A similar result holds if $h < 0$, so we have $m = f_x(a, b)$. The result $n = f_y(a, b)$ is found in a similar manner.

The previous example shows that if a function is differentiable at a point, it has partial derivatives there. Therefore, if any of the partial derivatives fail to exist, then the function cannot be differentiable. This is what happens in the following example of a cone.

Example 2 Consider the function $f(x, y) = \sqrt{x^2 + y^2}$. Is f differentiable at the origin?

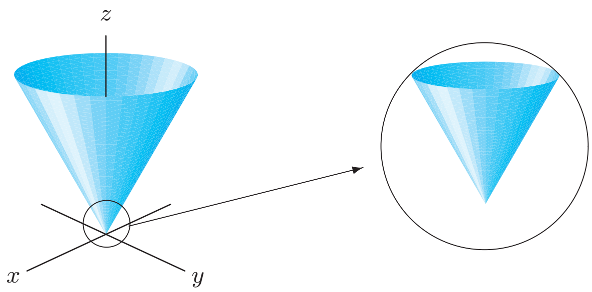

Figure I.36: The function $f(x, y) = \sqrt{x^2 + y^2}$ is not locally linear at $(0, 0)$: Zooming in around $(0, 0)$ does not make the graph look like a plane

Solution If we zoom in on the graph of the function $f(x, y) = \sqrt{x^2 + y^2}$ at the origin, as shown in Figure I.36, the sharp point remains; the graph never flattens out to look like a plane. Near its vertex, the graph does not look like it is well approximated (in any reasonable sense) by any plane.

Judging from the graph of f, we would not expect f to be differentiable at $(0, 0)$. Let us check this by trying to compute the partial derivatives of f at $(0, 0)$:

$$f_x(0, 0) = \lim_{h\to 0} \frac{f(h, 0) - f(0, 0)}{h} = \lim_{h\to 0} \frac{\sqrt{h^2 + 0} - 0}{h} = \lim_{h\to 0} \frac{|h|}{h}.$$

Since $|h|/h = \pm 1$, depending on whether h approaches 0 from the left or right, this limit does not exist and so neither does the partial derivative $f_x(0, 0)$. Thus, f cannot be differentiable at the origin. If it were, both of the partial derivatives, $f_x(0, 0)$ and $f_y(0, 0)$, would exist.

Alternatively, we could show directly that there is no linear approximation near $(0, 0)$ that satisfies the small relative error criterion for differentiability. Any plane passing through the point $(0, 0, 0)$ has the form $L(x, y) = mx + ny$ for some constants m and n. If $E(x, y) = f(x, y) - L(x, y)$, then

$$E(x, y) = \sqrt{x^2 + y^2} - mx - ny.$$

Then for f to be differentiable at the origin, we would need to show that

$$\lim_{\substack{h\to 0\\ k\to 0}} \frac{\sqrt{h^2 + k^2} - mh - nk}{\sqrt{h^2 + k^2}} = 0.$$

Taking $k = 0$ gives

$$\lim_{h\to 0} \frac{|h| - mh}{|h|} = 1 - m \lim_{h\to 0} \frac{h}{|h|}.$$

This limit exists only if $m = 0$ for the same reason as before. But then the value of the limit is 1 and not 0 as required. Thus, we again conclude f is not differentiable.

In Example 2 the partial derivatives f_x and f_y did not exist at the origin and this was sufficient to establish nondifferentiability there. We might expect that if both partial derivatives do exist, then f *is* differentiable. But the next example shows that this not necessarily true: the existence of both partial derivatives at a point is *not* sufficient to guarantee differentiability.

Example 3 Consider the function $f(x, y) = x^{1/3}y^{1/3}$. Show that the partial derivatives $f_x(0, 0)$ and $f_y(0, 0)$ exist, but that f is not differentiable at $(0, 0)$.

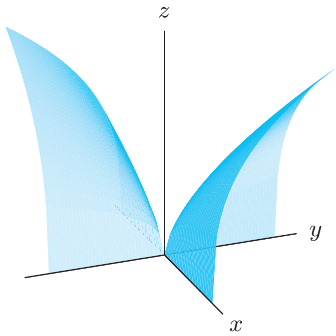

Figure I.37: Graph of $z = x^{1/3}y^{1/3}$ for $z \ge 0$

Solution

See Figure I.37 for the part of the graph of $z = x^{1/3}y^{1/3}$ when $z \ge 0$. We have $f(0,0) = 0$ and we compute the partial derivatives using the definition:

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h\to 0} \frac{0-0}{h} = 0,$$

and similarly

$$f_y(0,0) = 0.$$

So, if there did exist a linear approximation near the origin, it would have to be $L(x,y) = 0$. But we can show that this choice of $L(x,y)$ doesn't result in the small relative error that is required for differentiability. In fact, since $E(x,y) = f(x,y) - L(x,y) = f(x,y)$, we need to look at the limit

$$\lim_{\substack{h\to 0\\ k\to 0}} \frac{h^{1/3}k^{1/3}}{\sqrt{h^2+k^2}}.$$

If this limit exists, we get the same value no matter how h and k approach 0. Suppose we take $k = h > 0$. Then the limit becomes

$$\lim_{h\to 0} \frac{h^{1/3}h^{1/3}}{\sqrt{h^2+h^2}} = \lim_{h\to 0} \frac{h^{2/3}}{h\sqrt{2}} = \lim_{h\to 0} \frac{1}{h^{1/3}\sqrt{2}}.$$

But this limit does not exist, since small values for h will make the fraction arbitrarily large. So the only possible candidate for a linear approximation at the origin does not have a sufficiently small relative error. Thus, this function is *not* differentiable at the origin, even though the partial derivatives $f_x(0,0)$ and $f_y(0,0)$ exist. Figure I.37 confirms that near the origin the graph of $z = f(x,y)$ is not well approximated by any plane.

In summary,

- If a function is differentiable at a point, then both partial derivatives exist there.
- Having both partial derivatives at a point does not guarantee that a function is differentiable there.

Continuity and Differentiability

We know that differentiable functions of one variable are continuous. Similarly, it can be shown that if a function of two variables is differentiable at a point, then the function is continuous there.

In Example 3 the function f was continuous at the point where it was not differentiable. Example 4 shows that even if the partial derivatives of a function exist at a point, the function is not necessarily continuous at that point if it is not differentiable there.

Example 4 Suppose that f is the function of two variables defined by

$$f(x,y)=\begin{cases}\dfrac{xy}{x^2+y^2}, & (x,y)\neq(0,0),\\ 0, & (x,y)=(0,0).\end{cases}$$

Problem 4 on page 41 showed that $f(x,y)$ is not continuous at the origin. Show that the partial derivatives $f_x(0,0)$ and $f_y(0,0)$ exist. Could f be differentiable at $(0,0)$?

Solution From the definition of the partial derivative we see that

$$f_x(0,0)=\lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h}=\lim_{h\to 0}\left(\frac{1}{h}\cdot\frac{0}{h^2+0^2}\right)=\lim_{h\to 0}\frac{0}{h}=0,$$

and similarly

$$f_y(0,0)=0.$$

So, the partial derivatives $f_x(0,0)$ and $f_y(0,0)$ exist. However, f cannot be differentiable at the origin since it is not continuous there.

In summary,

- If a function is differentiable at a point, then it is continuous there.
- Having both partial derivatives at a point does not guarantee that a function is continuous there.

How Do We Know If a Function Is Differentiable?

Can we use partial derivatives to tell us if a function is differentiable? As we see from Examples 3 and 4, it is not enough that the partial derivatives exist. However, the following condition *does* guarantee differentiability:

If the partial derivatives, f_x and f_y, of a function f exist and are continuous on a small disk centered at the point (a,b), then f is differentiable at (a,b).

We will not prove this fact, although it provides a criterion for differentiability which is often simpler to use than the definition. It turns out that the requirement of continuous partial derivatives is more stringent than that of differentiability, so there exist differentiable functions which do not have continuous partial derivatives. However, most functions we encounter will have continuous partial derivatives. The class of functions with continuous partial derivatives is given the name C^1.

Example 5 Show that the function $f(x,y)=\ln(x^2+y^2)$ is differentiable everywhere in its domain.

Solution The domain of f is all of 2-space except for the origin. We shall show that f has continuous partial derivatives everywhere in its domain (that is, the function f is in C^1). The partial derivatives are

$$f_x=\frac{2x}{x^2+y^2}\quad\text{and}\quad f_y=\frac{2y}{x^2+y^2}.$$

Since each of f_x and f_y is the quotient of continuous functions, the partial derivatives are continuous everywhere except the origin (where the denominators are zero). Thus, f is differentiable everywhere in its domain.

Most functions built up from elementary functions have continuous partial derivatives, except perhaps at a few obvious points. Thus, in practice, we can often identify functions as being C^1 without explicitly computing the partial derivatives.

The Error in Linear and Quadratic Taylor Approximations

On page 809 of the textbook, we saw how to approximate a function $f(x, y)$ by Taylor polynomials. (The Taylor polynomial of degree 1 is the local linearization.) We now compare the magnitudes of the errors in the linear and quadratic approximations.

Let's return to the function $f(x, y) = \cos(2x+y)+\sin(x+y)$. The contour plots in Example 4 on page 810 of the textbook suggest that the quadratic approximation, $Q(x, y)$, is a better approximation to f than the linear approximation, $L(x, y)$. Consider approximations about the point $(0, 0)$. The errors in the linear and the quadratic approximations are defined as

$$E_L = f(x, y) - L(x, y) \qquad E_Q = f(x, y) - Q(x, y).$$

Table I.3 shows how the magnitudes of these errors, $|E_L|$ and $|E_Q|$, depend on the distance, $d(x, y) = \sqrt{x^2 + y^2}$, of the point (x, y) from $(0, 0)$. The values in Table I.3 suggest that, in this example,

$$E_L \text{ is proportional to } d^2 \quad \text{and} \quad E_Q \text{ is proportional to } d^3.$$

In general, the errors E_L and E_Q can be shown to be proportional to d^2 and d^3, respectively.

Table I.3 *Magnitude of the error in the linear and quadratic approximations to $f(x, y) = \cos(2x + y) + \sin(x + y)$*

Point, (x, y)	Distance, d	Error, $\|E_L\|$	Error, $\|E_Q\|$
$x = y = 0$	0	0	0
$x = y = 10^{-1}$	$1.4 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$4 \cdot 10^{-3}$
$x = y = 10^{-2}$	$1.4 \cdot 10^{-2}$	$5 \cdot 10^{-4}$	$4 \cdot 10^{-6}$
$x = y = 10^{-3}$	$1.4 \cdot 10^{-3}$	$5 \cdot 10^{-6}$	$4 \cdot 10^{-9}$
$x = y = 10^{-4}$	$1.4 \cdot 10^{-4}$	$5 \cdot 10^{-8}$	$4 \cdot 10^{-12}$

To use these approximations in practice, we need bounds on the magnitudes of the errors. If the distance between (x, y) and (a, b) is represented by $d(x, y) = \sqrt{(x - a)^2 + (y - b)^2}$, it can be shown that the following results hold:

Error Bound for Linear Approximation

Suppose $f(x, y)$ is a function with continuous second-order partial derivatives such that for $d(x, y) \leq d_0$,

$$|f_{xx}|, \ |f_{xy}|, \ |f_{yy}| \leq M_L.$$

Suppose

$$\begin{aligned} f(x, y) &= L(x, y) + E_L(x, y) \\ &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + E_L(x, y). \end{aligned}$$

Then we have

$$|E_L(x, y)| \leq 2M_L d(x, y)^2 \quad \text{for} \quad d(x, y) \leq d_0.$$

Note that the upper bound for the error term $E_L(x, y)$ has a form reminiscent of the second-order term in the Taylor formula for $f(x, y)$.

Error Bound for Quadratic Approximation

Suppose $f(x, y)$ is a function with continuous third-order partial derivatives such that for $d(x, y) \leq d_0$,

$$|f_{xxx}|,\ |f_{xxy}|,\ |f_{xyy}|,\ |f_{yyy}| \leq M_Q.$$

Suppose

$$\begin{aligned} f(x, y) &= Q(x, y) + E_Q(x, y) \\ &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &\quad + \frac{f_{xx}(a, b)}{2}(x - a)^2 + f_{xy}(a, b)(x - a)(y - b) + \frac{f_{yy}(a, b)}{2}(y - b)^2 + E_Q(x, y). \end{aligned}$$

Then we have

$$|E_Q(x, y)| \leq \frac{4}{3} M_Q d(x, y)^3 \quad \text{for} \quad d(x, y) \leq d_0.$$

Problem 15 shows how these error estimates and the coefficients (2 and $4/3$) are obtained. The important thing to notice is the fact that, for small d, the magnitude of E_L is much smaller than d and the magnitude of E_Q is much smaller than d^2. In other words we have the following result:

As $d(x, y) \to 0$:

$$\frac{E_L(x, y)}{d(x, y)} \to 0 \qquad \text{and} \qquad \frac{E_Q(x, y)}{(d(x, y))^2} \to 0.$$

This means that near the point (a, b), we can view the original function and the approximation as indistinguishable and behaving the same way.

Example 6 Suppose that the Taylor polynomial of degree 2 for f at $(0, 0)$ is $Q(x, y) = 5x^2 + 3y^2$. Suppose we are also told that

$$|f_{xxx}|, |f_{xxy}|, |f_{xyy}|, |f_{yyy}| \leq 9.$$

Notice that $Q(x, y) > 0$ for all (x, y) except $(0, 0)$. Show that, except at $(0, 0)$, we have

$$f(x, y) > 0 \quad \text{for all } (x, y) \text{ such that } \sqrt{x^2 + y^2} = d < 0.25.$$

Solution By the error bound for the Taylor polynomial of degree 2, we have

$$|E_Q(x, y)| = |f(x, y) - Q(x, y)| \leq \frac{4}{3}(9)d^3 = 12d^3$$

which can be written as

$$-12d^3 \leq f(x, y) - Q(x, y) \leq 12d^3.$$

Therefore we know that

$$Q(x, y) - 12d^3 \leq f(x, y).$$

Since $Q(x, y) = 5x^2 + 3y^2$, we have

$$5x^2 + 3y^2 - 12d^3 \leq f(x, y).$$

Since $5x^2 + 3y^2 \geq 3x^2 + 3y^2 = 3d^2$, we have

$$3d^2 - 12d^3 \leq f(x, y).$$

Now d^3 approaches 0 faster than d^2, so when d is small, we have

$$0 \leq 3d^2 - 12d^3 \leq f(x, y).$$

In fact, writing $3d^2 - 12d^3 = 3d^2(1 - 4d)$ shows that $d < 1/4$ ensures that $f(x, y) > 0$, except at $(0, 0)$ where $f = 0$. Thus, f has the same sign as Q for points near $(0, 0)$.

Problems for Section I

For the functions f in Problems 1–4 answer the following questions. Justify your answers.

(a) Use a computer to draw a contour diagram for f.
(b) Is f differentiable at all points $(x, y) \neq (0, 0)$?
(c) Do the partial derivatives f_x and f_y exist and are they continuous at all points $(x, y) \neq (0, 0)$?
(d) Is f differentiable at $(0, 0)$?
(e) Do the partial derivatives f_x and f_y exist and are they continuous at $(0, 0)$?

1. $f(x, y) = \begin{cases} \dfrac{x}{y} + \dfrac{y}{x}, & x \neq 0 \text{ and } y \neq 0, \\ 0, & x = 0 \text{ or } y = 0. \end{cases}$

2. $f(x, y) = \begin{cases} \dfrac{2xy}{(x^2 + y^2)^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$

3. $f(x, y) = \begin{cases} \dfrac{xy}{\sqrt{x^2 + y^2}}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$

4. $f(x, y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$

5. Consider the function

$$f(x, y) = \begin{cases} \dfrac{xy^2}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

(a) Use a computer to draw the contour diagram for f.
(b) Is f differentiable for $(x, y) \neq (0, 0)$?
(c) Show that $f_x(0, 0)$ and $f_y(0, 0)$ exist.
(d) Is f differentiable at $(0, 0)$?
(e) Suppose $x(t) = at$ and $y(t) = bt$, where a and b are constants, not both zero. If $g(t) = f(x(t), y(t))$, show that

$$g'(0) = \frac{ab^2}{a^2 + b^2}.$$

(f) Show that

$$f_x(0, 0)x'(0) + f_y(0, 0)y'(0) = 0.$$

Does the chain rule hold for the composite function $g(t)$ at $t = 0$? Explain.

(g) Show that the directional derivative $f_{\vec{u}}(0, 0)$ exists for each unit vector $\vec{u}$. Does this imply that f is differentiable at $(0, 0)$?

6. Consider the function

$$f(x, y) = \begin{cases} \dfrac{xy^2}{x^2 + y^4}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

(a) Use a computer to draw the contour diagram for f.
(b) Show that the directional derivative $f_{\vec{u}}(0, 0)$ exists for each unit vector $\vec{u}$.
(c) Is f continuous at $(0, 0)$? Is f differentiable at $(0, 0)$? Explain.

7. Consider the function $f(x, y) = \sqrt{|xy|}$.

(a) Use a computer to draw the contour diagram for f. Does the contour diagram look like that of a plane when we zoom in on the origin?
(b) Use a computer to draw the graph of f. Does the graph look like a plane when we zoom in on the origin?
(c) Is f differentiable for $(x, y) \neq (0, 0)$?
(d) Show that $f_x(0, 0)$ and $f_y(0, 0)$ exist.
(e) Is f differentiable at $(0, 0)$? [Hint: Consider the directional derivative $f_{\vec{u}}(0, 0)$ for $\vec{u} = (\vec{i} + \vec{j})/\sqrt{2}$.]

8. Suppose a function f is differentiable at the point (a, b). Show that f is continuous at (a, b).

9. Suppose $f(x, y)$ is a function such that $f_x(0, 0) = 0$ and $f_y(0, 0) = 0$, and $f_{\vec{u}}(0, 0) = 3$ for $\vec{u} = (\vec{i} + \vec{j})/\sqrt{2}$.

(a) Is f differentiable at $(0, 0)$? Explain.
(b) Give an example of a function f defined on 2-space which satisfies these conditions. [Hint: The function f does not have to be defined by a single formula valid over all of 2-space.]

10. Consider the following function:

$$f(x,y) = \begin{cases} \dfrac{xy(x^2-y^2)}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

The graph of f is shown in Figure I.38, and the contour diagram of f is shown in Figure I.39.

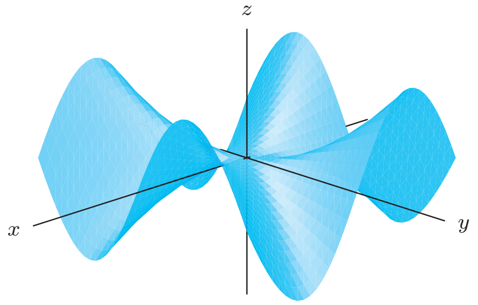

Figure I.38: Graph of $\dfrac{xy(x^2-y^2)}{x^2+y^2}$

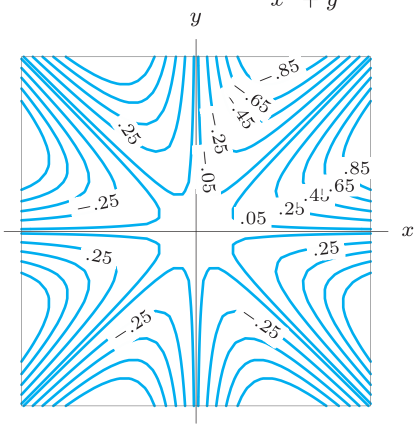

Figure I.39: Contour diagram of $\dfrac{xy(x^2-y^2)}{x^2+y^2}$

(a) Find $f_x(x,y)$ and $f_y(x,y)$ for $(x,y) \neq (0,0)$.
(b) Show that $f_x(0,0) = 0$ and $f_y(0,0) = 0$.
(c) Are the functions f_x and f_y continuous at $(0,0)$?
(d) Is f differentiable at $(0,0)$?

For Problems 11–14:

(a) Find the local linearization, $L(x,y)$, to the function $f(x,y)$ at the origin. Estimate the error $E_L(x,y) = f(x,y) - L(x,y)$ if $|x| \leq 0.1$ and $|y| \leq 0.1$.
(b) Find the degree 2 Taylor polynomial, $Q(x,y)$, for the function $f(x,y)$ at the origin. Estimate the error $E_Q(x,y) = f(x,y) - Q(x,y)$ if $|x| \leq 0.1$ and $|y| \leq 0.1$.
(c) Use a calculator to compute exactly $f(0.1, 0.1)$ and the errors $E_L(0.1,0.1)$ and $E_Q(0.1,0.1)$. How do these values compare with the errors predicted in parts (a) and (b)?

11. $f(x,y) = (\cos x)(\cos y)$

12. $f(x,y) = (e^x - x)\cos y$

13. $f(x,y) = e^{x+y}$

14. $f(x,y) = (x^2+y^2)e^{x+y}$

15. It is known that if the derivatives of a one-variable function, $g(t)$, satisfy

$$|g^{(n+1)}(t)| \leq K \quad \text{for } |t| \leq d_0,$$

then the error, E_n, in the n^{th} Taylor approximation, $P_n(x)$, is bounded as follows:

$$|E_n| = |g(t) - P_n(t)| \leq \frac{K}{(n+1)!}|t|^{n+1} \quad \text{for } |t| \leq d_0.$$

In this problem, we use this result for $g(t)$ to get the error bounds for the linear and quadratic Taylor approximations to $f(x,y)$. For a particular function $f(x,y)$, let $x = ht$ and $y = kt$ for fixed h and k, and define $g(t)$ as follows:

$$g(t) = f(ht, kt) \quad \text{for } 0 \leq t \leq 1.$$

(a) Calculate $g'(t)$, $g''(t)$, and $g'''(t)$ using the chain rule.
(b) Show that $L(ht,kt) = P_1(t)$ and that $Q(ht,kt) = P_2(t)$, where L is the linear approximation to f at $(0,0)$ and Q is the Taylor polynomial of degree 2 for f at $(0,0)$.
(c) What is the relation between $E_L = f(x,y) - L(x,y)$ and E_1? What is the relation between $E_Q = f(x,y) - Q(x,y)$ and E_2?
(d) Assuming that the second and third-order partial derivatives of f are bounded for $d(x,y) \leq d_0$, show that $|E_L|$ and $|E_Q|$ are bounded as on page 47.

J EXISTENCE OF GLOBAL EXTREMA FOR FUNCTIONS OF MANY VARIABLES

Under what circumstances does a function of two variables have a global maximum or minimum? The next example shows that a function may have both a global maximum and a global minimum on a region, or just one, or neither.

Example 1 Investigate the global maxima and minima of the following functions:

(a) $h(x, y) = 1 + x^2 + y^2$ on the disk $x^2 + y^2 \le 1$.
(b) $f(x, y) = x^2 - 2x + y^2 - 4y + 5$ on the xy-plane.
(c) $g(x, y) = x^2 - y^2$ on the xy-plane.

Solution

(a) The graph of $h(x, y) = 1 + x^2 + y^2$ is a bowl shaped paraboloid with a global minimum of 1 at $(0, 0)$, and a global maximum of 2 on the edge of the region, $x^2 + y^2 = 1$.
(b) The graph of f in Figure **??** on page **??** of the textbook shows that f has a global minimum at the point $(1, 2)$ and no global maximum (because the value of f increases without bound as $x \to \infty, y \to \infty$).
(c) The graph of g in Figure **??** on page **??** of the textbook shows that g has no global maximum because $g(x, y) \to \infty$ as $x \to \infty$ if y is constant. Similarly, g has no global minimum because $g(x, y) \to -\infty$ as $y \to \infty$ if x is constant.

There are, however, conditions that guarantee that a function has a global maximum and minimum. For $h(x)$, a function of one variable, the function must be continuous on a closed interval $a \le x \le b$. If h is continuous on a non-closed interval, such as $a \le x < b$ or $a < x < b$, or on an interval which is not bounded, such as $a < x < \infty$, then h need not have a maximum or minimum value. What is the situation for functions of two variables? As it turns out, a similar result is true for continuous functions defined on regions which are closed and bounded, analogous to the closed and bounded interval $a \le x \le b$. In everyday language we say

- A **closed** region is one which contains its boundary;
- A **bounded** region is one which does not stretch to infinity in any direction.

More precise definitions are as follows. Suppose R is a region in 2-space. A point (x_0, y_0) is a *boundary point* of R if, for every $r > 0$, the disk $(x - x_0)^2 + (y - y_0)^2 < r^2$ with center (x_0, y_0) and radius r contains both points which are in R and points which are not in R. See Figure J.40. A point (x_0, y_0) can be a boundary point of the region R without actually belonging to R. A point (x_0, y_0) in R is an *interior point* if it is not a boundary point; thus, for small enough $r > 0$, the disk of radius r centered at (x_0, y_0) lies entirely in the region R. See Figure J.41. The collection of all the boundary points is the *boundary of* R and the collection of all the interior points is the *interior* of R. The region R is *closed* if it contains its boundary, while it is *open* if every point in R is an interior point.

A region R in 2-space is *bounded* if the distance between every point (x, y) in R and the origin is less than or equal to some constant number K. Closed and bounded regions in 3-space are defined in the same way.

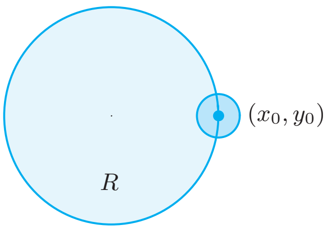

Figure J.40: Boundary point (x_0, y_0) of R

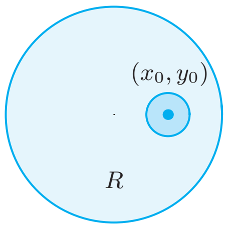

Figure J.41: Interior point (x_0, y_0) of R

Example 2 (a) The square $-1 \le x \le 1$, $-1 \le y \le 1$ is closed and bounded.
(b) The first quadrant $x \ge 0$, $y \ge 0$ is closed but is not bounded.
(c) The disk $x^2 + y^2 < 1$ is open and bounded, but is not closed.
(d) The half-plane $y > 0$ is open, but is neither closed nor bounded.

The reason that closed and bounded regions are useful is the following result[8]:

> If f is a continuous function on a closed and bounded region R, then f has a global maximum at some point (x_0, y_0) in R and a global minimum at some point (x_1, y_1) in R.

The result is also true for functions of three or more variables.

If f is not continuous or the region R is not closed and bounded, there is no guarantee that f will achieve a global maximum or global minimum on R. In Example 1, the function g is continuous but does not achieve a global maximum or minimum in 2-space, a region which is closed but not bounded. The following example illustrates what can go wrong when the region is bounded but not closed.

Example 3 Does the following function have a global maximum or minimum on the region R given by $0 < x^2 + y^2 \le 1$?

$$f(x, y) = \frac{1}{x^2 + y^2}$$

Solution The region R is bounded, but it is not closed since it does not contain the boundary point $(0, 0)$. We see from the graph of $z = f(x, y)$ in Figure J.42 that f has a global minimum on the circle $x^2 + y^2 = 1$. However, $f(x, y) \to \infty$ as $(x, y) \to (0, 0)$, so f has no global maximum.

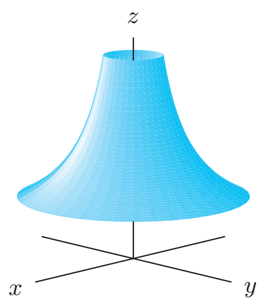

Figure J.42: Graph showing $f(x, y) = \frac{1}{x^2+y^2}$ has no global maximum on $0 < x^2 + y^2 \le 1$

K CHANGE OF VARIABLES IN A MULTIPLE INTEGRAL

In Chapter 16 we used polar, cylindrical, and spherical coordinates to simplify iterated integrals. In this section, we discuss more general changes of variable. In the process, we will see where the extra factor of r comes from when we change from Cartesian to polar coordinates and the factor $\rho^2 \sin \phi$ when we change from Cartesian to spherical coordinates.

[8]For a proof, see W. Rudin, *Principles of Mathematical Analysis*, 2nd ed., p. 89, (New York: McGraw-Hill, 1976)

Polar Change of Variables Revisited

Consider the integral $\int_R (x+y)\,dA$ where R is the region in the first quadrant bounded by the circle $x^2+y^2=16$ and the x and y-axes. Writing the integral in Cartesian and polar coordinates we have

$$\int_R (x+y)\,dA = \int_0^4 \int_0^{\sqrt{16-x^2}} (x+y)\,dy\,dx = \int_0^{\pi/2} \int_0^4 (r\cos\theta + r\sin\theta) r\,dr d\theta.$$

This is an integral over the rectangle in the $r\theta$-space given by $0 \le r \le 4$, $0 \le \theta \le \pi/2$. The conversion from polar to Cartesian coordinates changes this rectangle into a quarter-disk. Figure K.43 shows how a typical rectangle (shaded) in the $r\theta$-plane with sides of length Δr and $\Delta\theta$ corresponds to a curved rectangle in the xy-plane with sides of length Δr and $r\Delta\theta$. The extra r is needed because the correspondence between r, θ and x, y not only curves the lines $r = 1, 2, 3 \ldots$ into circles, it also stretches those lines around larger and larger circles.

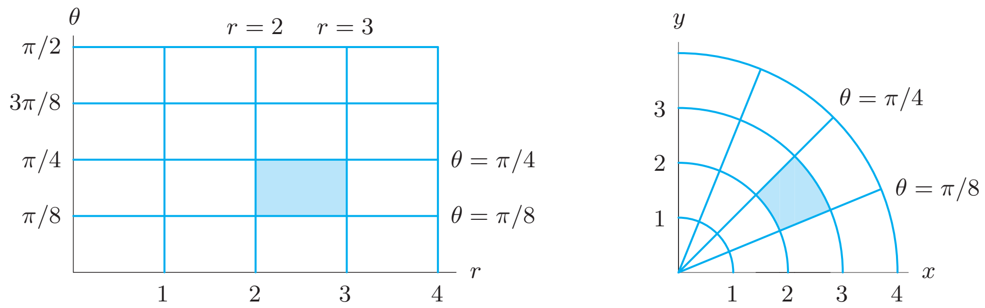

Figure K.43: A grid in the $r\theta$-plane and the corresponding curved grid in the xy-plane

General Change of Variables

We now consider a general change of variable, where x, y coordinates are related to s, t coordinates by the differentiable functions

$$x = x(s,t) \quad y = y(s,t).$$

Just as a rectangular region in the $r\theta$-plane corresponds to a circular region in the xy-plane, a rectangular region, T, in the st-plane corresponds to a curved region, R, in the xy-plane. We assume that the change of coordinates is one-to-one, that is, that each point R corresponds to one point in T.

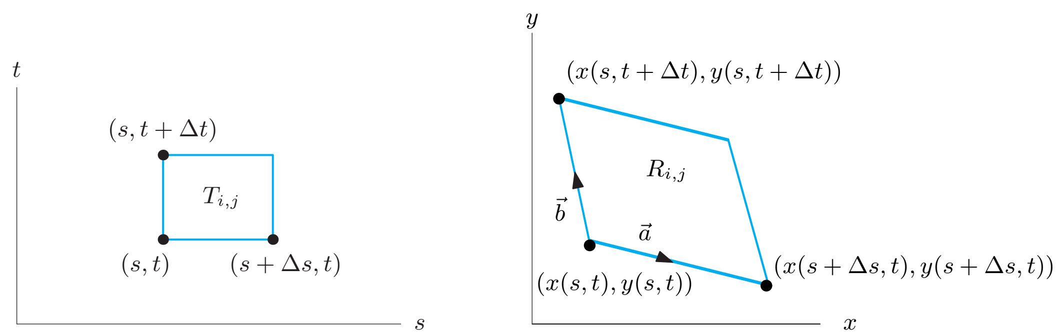

Figure K.44: A small rectangle $T_{i,j}$ in the st-plane and the corresponding region $R_{i,j}$ of the xy-plane

We divide T into small rectangles $T_{i,j}$ with sides of length Δs and Δt. (See Figure K.44.) The corresponding piece $R_{i,j}$ of the xy-plane is a quadrilateral with curved sides. If we choose Δs and Δt very small, then by local linearity, $R_{i,j}$ is approximately a parallelogram.

Recall from Chapter 13 that the area of the parallelogram with sides $\vec{a}$ and $\vec{b}$ is $\|\vec{a} \times \vec{b}\|$. Thus, we need to find the sides of $R_{i,j}$ as vectors. The side of $R_{i,j}$ corresponding to the bottom side of $T_{i,j}$ has endpoints $(x(s,t), y(s,t))$ and $(x(s+\Delta s, t), y(s+\Delta s, t))$, so in vector form that side is

$$\vec{a} = (x(s+\Delta s,t)-x(s,t))\vec{i} + (y(s+\Delta s,t)-y(s,t))\vec{j} + 0\vec{k} \approx \left(\frac{\partial x}{\partial s}\Delta s\right)\vec{i} + \left(\frac{\partial y}{\partial s}\Delta s\right)\vec{j} + 0\vec{k}\,.$$

Similarly, the side of $R_{i,j}$ corresponding to the left edge of $T_{i,j}$ is given by

$$\vec{b} \approx \left(\frac{\partial x}{\partial t}\Delta t\right)\vec{i} + \left(\frac{\partial y}{\partial t}\Delta t\right)\vec{j} + 0\vec{k}\,.$$

Computing the cross product, we get

$$\begin{aligned}\text{Area } R_{i,j} \approx \|\vec{a} \times \vec{b}\| &\approx \left|\left(\frac{\partial x}{\partial s}\Delta s\right)\left(\frac{\partial y}{\partial t}\Delta t\right) - \left(\frac{\partial x}{\partial t}\Delta t\right)\left(\frac{\partial y}{\partial s}\Delta s\right)\right| \\ &= \left|\frac{\partial x}{\partial s}\cdot\frac{\partial y}{\partial t} - \frac{\partial x}{\partial t}\cdot\frac{\partial y}{\partial s}\right| \Delta s \Delta t.\end{aligned}$$

Using determinant notation, we define the *Jacobian*, $\dfrac{\partial(x,y)}{\partial(s,t)}$, as follows

$$\frac{\partial(x,y)}{\partial(s,t)} = \frac{\partial x}{\partial s}\cdot\frac{\partial y}{\partial t} - \frac{\partial x}{\partial t}\cdot\frac{\partial y}{\partial s} = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \end{vmatrix}.$$

Thus, we can write

$$\text{Area } R_{i,j} \approx \left|\frac{\partial(x,y)}{\partial(s,t)}\right| \Delta s\, \Delta t.$$

To compute $\int_R f(x,y)\,dA$, where f is a continuous function, we look at the Riemann sum obtained by dividing the region R into the small curved regions $R_{i,j}$, giving

$$\int_R f(x,y)\,dA \approx \sum_{i,j} f(x_i,y_j)\cdot \text{Area of } R_{i,j} \approx \sum_{i,j} f(x_i,y_j)\left|\frac{\partial(x,y)}{\partial(s,t)}\right| \Delta s\, \Delta t.$$

Each point (x_i, y_j) corresponds to a point (s_i, t_j), so the sum can be written in terms of s and t:

$$\sum_{i,j} f(x(s_i,t_j), y(s_i,t_j))\left|\frac{\partial(x,y)}{\partial(s,t)}\right| \Delta s\, \Delta t.$$

This is a Riemann sum in terms of s and t, so as Δs and Δt approach 0, we get

$$\int_R f(x,y)\,dA = \int_T f(x(s,t),y(s,t))\left|\frac{\partial(x,y)}{\partial(s,t)}\right| ds\,dt.$$

To convert an integral from x, y to s, t coordinates we make three changes:

1. Substitute for x and y in the integrand in terms of s and t.
2. Change the xy region R into an st region T.
3. Introduce the absolute value of the Jacobian, $\left|\dfrac{\partial(x,y)}{\partial(s,t)}\right|$, representing the change in the area element.

Example 1 Verify that the Jacobian $\dfrac{\partial(x,y)}{\partial(r,\theta)} = r$ for polar coordinates $x = r\cos\theta,\ y = r\sin\theta$.

Solution
$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Example 2 Find the area of the ellipse $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$.

Solution Let $x = as,\ y = bt$. Then the ellipse $x^2/a^2 + y^2/b^2 = 1$ in the xy-plane corresponds to the circle $s^2 + t^2 = 1$ in the st-plane. The Jacobian is $\begin{vmatrix} a & 0 \\ 0 & b \end{vmatrix} = ab$. Thus, if we let R be the ellipse in the xy-plane and T the unit circle in the st-plane, we get

$$\text{Area of } xy\text{-ellipse} = \int_R 1\,dA = \int_T 1ab\,ds\,dt = ab\int_T ds\,dt = ab \cdot \ \text{Area of } st\text{-circle} = \pi ab.$$

Change of Variables in Triple Integrals

For triple integrals, there is a similar formula. Suppose the differentiable functions

$$x = x(s,t,u), \quad y = y(s,t,u), \quad z = z(s,t,u)$$

define a change of variables from a region S in stu-space to a region W in xyz-space. Then, the Jacobian of this change of variables is given by the determinant

$$\frac{\partial(x,y,z)}{\partial(s,t,u)} = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} & \frac{\partial z}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} & \frac{\partial z}{\partial t} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \end{vmatrix}.$$

Just as the Jacobian in two dimensions gives us the change in the area element, the Jacobian in three dimensions represents the change in the volume element. Thus, we have

$$\int_W f(x,y,z)\,dx\,dy\,dz = \int_S f(x(s,t,u), y(s,t,u), z(s,t,u)) \left|\frac{\partial(x,y,z)}{\partial(s,t,u)}\right| ds\,dt\,du.$$

Problem 3 at the end of this section asks you to verify that the Jacobian for the change of variables for spherical coordinates is $\rho^2\sin\phi$. The next example generalizes Example 2 to ellipsoids.

Example 3 Find the volume of the ellipsoid $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$.

Solution Let $x = as,\ y = bt,\ z = cu$. The Jacobian is computed to be abc. The xyz-ellipsoid corresponds to the stu-sphere $s^2 + t^2 + u^2 = 1$. Thus, as in Example 2,

$$\text{Volume of } xyz\text{-ellipsoid} = abc \cdot \text{Volume of } stu\text{-sphere} = abc\frac{4}{3}\pi = \frac{4}{3}\pi abc.$$

Problems for Section K

1. Find the region R in the xy-plane corresponding to the region $T = \{(s,t)\,|\,0 \le s \le 3,\ 0 \le t \le 2\}$ under the change of variables $x = 2s - 3t,\ y = s - 2t$. Check that

$$\int_R dx\,dy = \int_T \left|\frac{\partial(x,y)}{\partial(s,t)}\right| ds\,dt.$$

2. Find the region R in the xy-plane corresponding to the region $T = \{(s,t)\,|\,0 \le s \le 2,\ s \le t \le 2\}$ under the change of variables $x = s^2,\ y = t$. Check that

$$\int_R dx\,dy = \int_T \left|\frac{\partial(x,y)}{\partial(s,t)}\right| ds\,dt.$$

3. Compute the Jacobian for the change of variables into spherical coordinates:

$$x = \rho\sin\phi\cos\theta, \quad y = \rho\sin\phi\sin\theta, \quad z = \rho\cos\phi.$$

4. For the change of variables $x = 3s - 4t,\ y = 5s + 2t$, show that

$$\frac{\partial(x,y)}{\partial(s,t)} \cdot \frac{\partial(s,t)}{\partial(x,y)} = 1$$

5. Use the change of variables $x = 2s + t,\ y = s - t$ to compute the integral $\int_R (x+y)\,dA$, where R is the parallelogram formed by $(0,0)$, $(3,-3)$, $(5,-2)$, and $(2,1)$.

6. Use the change of variables $x = \frac{1}{2}s,\ y = \frac{1}{3}t$ to compute the integral $\int_R (x^2 + y^2)\,dA$, where R is the region bounded by the curve $4x^2 + 9y^2 = 36$.

7. Use the change of variables $s = xy,\ t = xy^2$ to compute $\int_R xy^2\,dA$, where R is the region bounded by $xy = 1,\ xy = 4,\ xy^2 = 1,\ xy^2 = 4$.

8. Evaluate the integral $\displaystyle\int_R \cos\left(\frac{x-y}{x+y}\right) dxdy$ where R is the triangle bounded by $x + y = 1$, $x = 0$, and $y = 0$.

9. Find the area of the metal frames with one or four cutouts shown in Figure K.45. Start with Cartesian coordinates x, y aligned along one side. Consider slanted coordinates $u = x - y$, $v = y$ in which the frame is "straightened". [Hint: First describe the shape of the cut-out in the uv-plane; second, calculate its area in the uv-plane; third, using Jacobians, calculate its area in the xy-plane.]

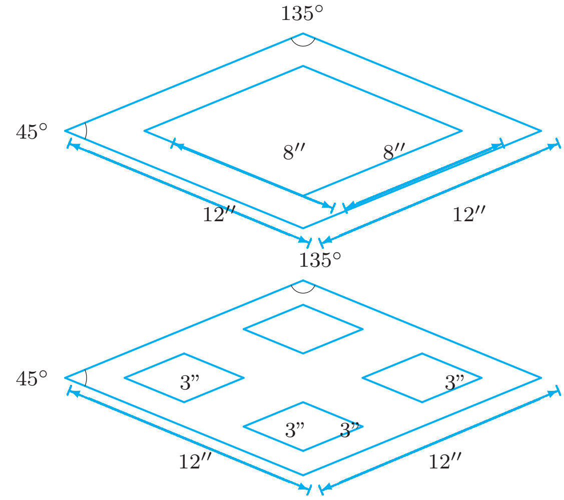

Figure K.45

10. A river follows the path $y = f(x)$ where x, y are in kilometers. Near the sea, it widens into a lagoon, then narrows again at its mouth. See Figure K.46. At the point (x,y), the depth, $d(x,y)$, of the lagoon is given by

$$d(x,y) = 40 - 160(y - f(x))^2 - 40x^2 \text{ meters.}$$

The lagoon itself is described by $d(x,y) \ge 0$. What is the volume of the lagoon in cubic meters? [Hint: Use new coordinates $u = x/2$, $v = y - f(x)$ and Jacobians.]

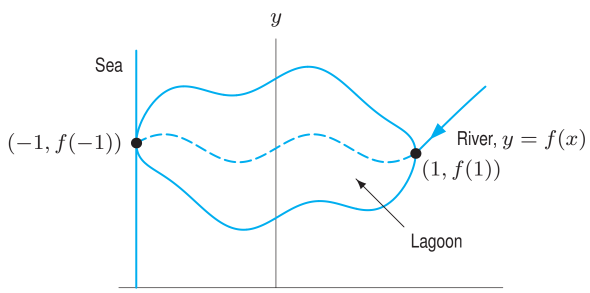

Figure K.46

L PROOF OF GREEN'S THEOREM

In this section we will give a proof of Green's Theorem based on the change of variables formula for double integrals. Assume the vector field $\vec{F}$ is given in components by

$$\vec{F}(x,y) = F_1(x,y)\vec{i} + F_2(x,y)\vec{j}.$$

Proof for Rectangles

We prove Green's Theorem first when R is a rectangular region, as shown in Figure L.47. The line integral in Green's theorem can be written as

$$\begin{aligned}\int_C \vec{F}\cdot d\vec{r} &= \int_{C_1} \vec{F}\cdot d\vec{r} + \int_{C_2} \vec{F}\cdot d\vec{r} + \int_{C_3} \vec{F}\cdot d\vec{r} + \int_{C_4} \vec{F}\cdot d\vec{r}\\ &= \int_a^b F_1(x,c)\,dx + \int_c^d F_2(b,y)\,dy - \int_a^b F_1(x,d)\,dx - \int_c^d F_2(a,y)\,dy\\ &= \int_c^d (F_2(b,y) - F_2(a,y))\,dy + \int_a^b (-F_1(x,d) + F_1(x,c))\,dx.\end{aligned}$$

On the other hand, the double integral in Green's theorem can be written as an iterated integral. We evaluate the inside integral using the Fundamental Theorem of Calculus.

$$\begin{aligned}\int_R \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) dx\,dy &= \int_R \frac{\partial F_2}{\partial x}\,dx\,dy + \int_R -\frac{\partial F_1}{\partial y}\,dx\,dy\\ &= \int_c^d \int_a^b \frac{\partial F_2}{\partial x}\,dx\,dy + \int_a^b \int_c^d -\frac{\partial F_1}{\partial y}\,dy\,dx\\ &= \int_c^d (F_2(b,y) - F_2(a,y))\,dy + \int_a^b (-F_1(x,d) + F_1(x,c))\,dx.\end{aligned}$$

Since the line integral and the double integral are equal, we have proved Green's theorem for rectangles.

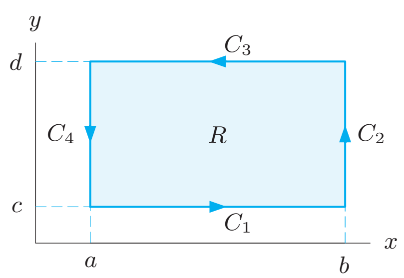

Figure L.47: A rectangular region R with boundary C broken into C_1, C_2, C_3, and C_4

Proof for Regions Parameterized by Rectangles

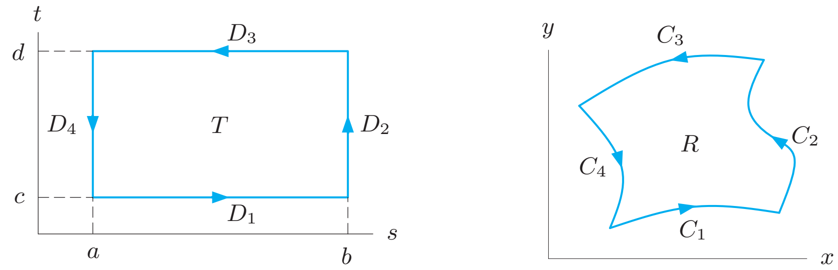

Figure L.48: A curved region R in the xy-plane corresponding to a rectangular region T in the st-plane

Now we prove Green's Theorem for a region R which can be transformed into a rectangular region. Suppose we have a smooth change of coordinates

$$x = x(s,t), \qquad y = y(s,t).$$

Consider a curved region R in the xy-plane corresponding to a rectangular region T in the st-plane, as in Figure L.48. We suppose that the change of coordinates is one-to-one on the interior of T.

We prove Green's theorem for R using Green's theorem for T and the change of variables formula for double integrals given on page **??**. First we express the line integral around C

$$\int_C \vec{F} \cdot d\vec{r},$$

as a line integral in the st-plane around the rectangle $D = D_1 + D_2 + D_3 + D_4$. In vector notation, the change of coordinates is

$$\vec{r} = \vec{r}(s,t) = x(s,t)\vec{i} + y(s,t)\vec{j}$$

and so

$$\vec{F} \cdot d\vec{r} = \vec{F}(\vec{r}(s,t)) \cdot \frac{\partial \vec{r}}{\partial s}\, ds + \vec{F}(\vec{r}(s,t)) \cdot \frac{\partial \vec{r}}{\partial t}\, dt.$$

We define a vector field $\vec{G}$ on the st-plane with components

$$G_1 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \quad \text{and} \quad G_2 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial t}.$$

Then, if $\vec{u}$ is the position vector of a point in the st-plane, we have $\vec{F} \cdot d\vec{r} = G_1\, ds + G_2\, dt = \vec{G} \cdot d\vec{u}$. Problem 5 at the end of this section asks you to show that the formula for line integrals along parameterized paths leads to the following result:

$$\int_C \vec{F} \cdot d\vec{r} = \int_D \vec{G} \cdot d\vec{u}.$$

In addition, using the product rule and chain rule we can show that

$$\frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t} = \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \end{vmatrix}.$$

(See Problem 6 at the end of this section.) Hence, by the change of variables formula for double integrals on page **??**,

$$\int_R \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) dx\, dy = \int_T \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \end{vmatrix} ds\, dt = \int_T \left(\frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t}\right) ds\, dt.$$

Thus we have shown that

$$\int_C \vec{F} \cdot d\vec{r} = \int_D \vec{G} \cdot d\vec{u}$$

and that

$$\int_R \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx\, dy = \int_T \left(\frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t} \right) ds\, dt.$$

The integrals on the right are equal, by Green's Theorem for rectangles; hence the integrals on the left are equal as well, which is Green's Theorem for the region R.

Pasting Regions Together

Lastly we show that Green's Theorem holds for a region formed by pasting together regions which can be transformed into rectangles. Figure L.49 shows two regions R_1 and R_2 that fit together to form a region R. We break the boundary of R into C_1, the part shared with R_1, and C_2, the part shared with R_2. We let C be the part of the the boundary of R_1 which it shares with R_2. So

$$\text{Boundary of } R = C_1 + C_2, \quad \text{Boundary of } R_1 = C_1 + C, \quad \text{Boundary of } R_2 = C_2 + (-C).$$

Note that when the curve C is considered as part of the boundary of R_2, it receives the opposite orientation from the one it receives as the boundary of R_1. Thus

$$\begin{aligned}
\int_{\text{Boundary of } R_1} \vec{F} \cdot d\vec{r} + \int_{\text{Boundary of } R_2} \vec{F} \cdot d\vec{r} &= \int_{C_1+C} \vec{F} \cdot d\vec{r} + \int_{C_2+(-C)} \vec{F} \cdot d\vec{r} \\
&= \int_{C_1} \vec{F} \cdot d\vec{r} + \int_{C} \vec{F} \cdot d\vec{r} + \int_{C_2} \vec{F} \cdot d\vec{r} - \int_{C} \vec{F} \cdot d\vec{r} \\
&= \int_{C_1} \vec{F} \cdot d\vec{r} + \int_{C_2} \vec{F} \cdot d\vec{r} \\
&= \int_{\text{Boundary of } R} \vec{F} \cdot d\vec{r}\,.
\end{aligned}$$

So, applying Green's Theorem for R_1 and R_2, we get

$$\begin{aligned}
\int_R \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx\, dy &= \int_{R_1} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx\, dy + \int_{R_2} \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx\, dy \\
&= \int_{\text{Boundary of } R_1} \vec{F} \cdot d\vec{r} + \int_{\text{Boundary of } R_2} \vec{F} \cdot d\vec{r} \\
&= \int_{\text{Boundary of } R} \vec{F} \cdot d\vec{r}\,,
\end{aligned}$$

which is Green's Theorem for R. Thus, we have proved Green's Theorem for any region formed by pasting together regions that are smoothly parameterized by rectangles.

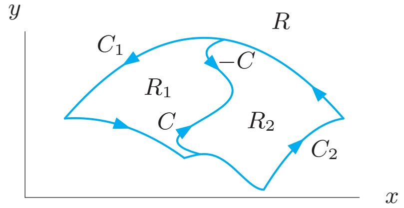

Figure L.49: Two regions R_1 and R_2 pasted together to form a region R

Example 1 Let R be the annulus (ring) centered at the origin with inner radius 1 and outer radius 2. Using polar coordinates, show that the proof of Green's Theorem applies to R. See Figure L.50.

Solution In polar coordinates, $x = r\cos t$ and $y = r\sin t$, the annulus corresponds to the rectangle in the rt-plane $1 \le r \le 2$, $0 \le t \le 2\pi$. The sides $t = 0$ and $t = 2\pi$ are pasted together in the xy-plane along the x-axis; the other two sides become the inner and outer circles of the annulus. Thus R is formed by pasting the ends of a rectangle together.

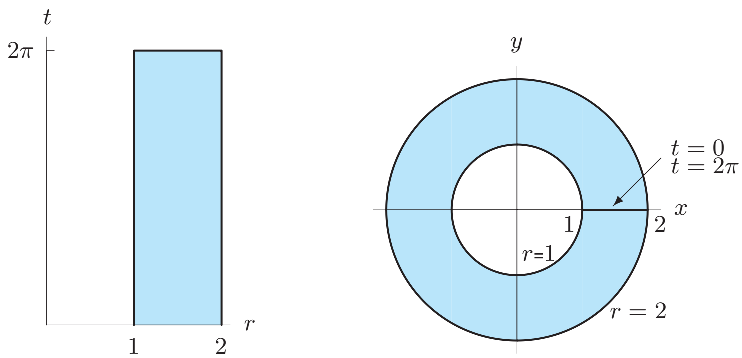

Figure L.50: The annulus R in the xy-plane and the corresponding rectangle $1 \le r \le 2$, $0 \le t \le 2\pi$ in the rt-plane

Problems for Section L

1. Let R be the annulus centered at $(-1, 2)$ with inner radius 2 and outer radius 3. Show that R can be parameterized by a rectangle.

2. Let R be the region under the first arc of the graph of the sine function. Show that R can be parameterized by a rectangle.

3. Let $f(x)$ and $g(x)$ be two smooth functions, and suppose that $f(x) \le g(x)$ for $a \le x \le b$. Let R be the region $f(x) \le y \le g(x)$, $a \le x \le b$.
   (a) Sketch an example of such a region.
   (b) For a constant x_0, parameterize the vertical line segment in R where $x = x_0$. Choose your parameterization so that the parameter starts at 0 and ends at 1.
   (c) By putting together the parameterizations in part (b) for different values of x_0, show that R can be parameterized by a rectangle.

4. Let $f(y)$ and $g(y)$ be two smooth functions, and suppose that $f(y) \le g(y)$ for $c \le y \le d$. Let R be the region $f(y) \le x \le g(y)$, $c \le y \le d$.
   (a) Sketch an example of such a region.
   (b) For a constant y_0, parameterize the horizontal line segment in R where $y = y_0$. Choose your parameterization so that the parameter starts at 0 and ends at 1.
   (c) By putting together the parameterizations in part (b) for different values of y_0, show that R can be parameterized by a rectangle.

5. Use the formula for calculating line integrals by parameterization to prove the statement on page 58:
$$\int_C \vec{F} \cdot d\vec{r} = \int_D \vec{G} \cdot d\vec{u}\,.$$

6. Use the product rule and the chain rule to prove the formula on page 58:
$$\frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t} = \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \end{vmatrix}.$$

M PROOF OF THE DIVERGENCE THEOREM AND STOKES' THEOREM

In this section we give proofs of the Divergence Theorem and Stokes' Theorem using the definitions in Cartesian coordinates.

Proof of the Divergence Theorem

For the Divergence Theorem, we use the same approach as we used for Green's Theorem; first prove the theorem for rectangular regions, then use the change of variables formula to prove it for regions parameterized by rectangular regions, and finally paste such regions together to form general regions.

Proof for Rectangular Solids with Sides Parallel to the Axes

Consider a smooth vector field $\vec{F}$ defined on the rectangular solid V: $a \le x \le b$, $c \le y \le d$, $e \le z \le f$. (See Figure M.51). We start by computing the flux of $\vec{F}$ through the two faces of V perpendicular to the x-axis, A_1 and A_2, both oriented outward:

$$\begin{aligned}\int_{A_1} \vec{F} \cdot d\vec{A} + \int_{A_2} \vec{F} \cdot d\vec{A} &= -\int_e^f \int_c^d F_1(a,y,z)\,dy\,dz + \int_e^f \int_c^d F_1(b,y,z)\,dy\,dz \\ &= \int_e^f \int_c^d (F_1(b,y,z) - F_1(a,y,z))\,dy\,dz.\end{aligned}$$

By the Fundamental Theorem of Calculus,

$$F_1(b,y,z) - F_1(a,y,z) = \int_a^b \frac{\partial F_1}{\partial x}\,dx,$$

so

$$\int_{A_1} \vec{F} \cdot d\vec{A} + \int_{A_2} \vec{F} \cdot d\vec{A} = \int_e^f \int_c^d \int_a^b \frac{\partial F_1}{\partial x}\,dx\,dy\,dz = \int_V \frac{\partial F_1}{\partial x}\,dV.$$

By a similar argument, we can show

$$\int_{A_3} \vec{F} \cdot d\vec{A} + \int_{A_4} \vec{F} \cdot d\vec{A} = \int_V \frac{\partial F_2}{\partial y}\,dV \quad \text{and} \quad \int_{A_5} \vec{F} \cdot d\vec{A} + \int_{A_6} \vec{F} \cdot d\vec{A} = \int_V \frac{\partial F_3}{\partial z}\,dV.$$

Adding these, we get

$$\int_A \vec{F} \cdot d\vec{A} = \int_V \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right) dV = \int_V \operatorname{div} \vec{F}\,dV.$$

This is the Divergence Theorem for the region V.

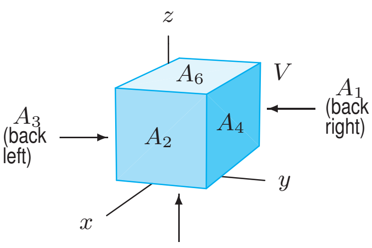

Figure M.51: Rectangular solid V in xyz-space

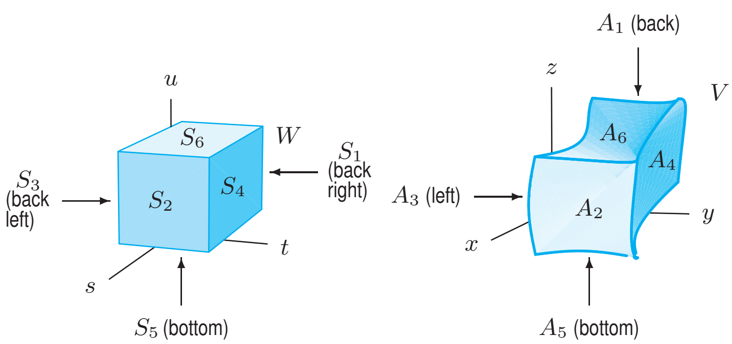

Figure M.52: A rectangular solid W in stu-space and the corresponding curved solid V in xyz-space

Proof for Regions Parameterized by Rectangular Solids

Now suppose we have a smooth change of coordinates

$$x = x(s,t,u), \qquad y = y(s,t,u), \quad z = z(s,t,u).$$

Consider a curved solid V in xyz-space corresponding to a rectangular solid W in stu-space. See Figure M.52. We suppose that the change of coordinates is one-to-one on the interior of W, and that its Jacobian determinant is positive on W. We prove the Divergence Theorem for V using the Divergence Theorem for W.

Let A be the boundary of V. To prove the Divergence Theorem for V, we must show that

$$\int_A \vec{F} \cdot d\vec{A} = \int_V \operatorname{div} \vec{F}\, dV.$$

First we express the flux through A as a flux integral in stu-space over S, the boundary of the rectangular region W. In vector notation the change of coordinates is

$$\vec{r} = \vec{r}(s,t,u) = x(s,t,u)\vec{i} + y(s,t,u)\vec{j} + z(s,t,u)\vec{k}.$$

The face A_1 of V is parameterized by

$$\vec{r} = \vec{r}(a,t,u), \qquad c \le t \le d, \quad e \le u \le f,$$

so on this face

$$d\vec{A} = \pm\frac{\partial \vec{r}}{\partial t} \times \frac{\partial \vec{r}}{\partial u}.$$

In fact, in order to make $d\vec{A}$ point outward, we must choose the negative sign. (Problem 3 on page 66 shows how this follows from the fact that the Jacobian determinant is positive.) Thus, if S_1 is the face $s = a$ of W,

$$\int_{A_1} \vec{F} \cdot d\vec{A} = -\int_{S_1} \vec{F} \cdot \frac{\partial \vec{r}}{\partial t} \times \frac{\partial \vec{r}}{\partial u}\, dt\, du,$$

The outward pointing area element on S_1 is $d\vec{S} = -\vec{i}\, dt\, du$. Therefore, if we choose a vector field $\vec{G}$ on stu-space whose component in the s-direction is

$$G_1 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial t} \times \frac{\partial \vec{r}}{\partial u},$$

we have

$$\int_{A_1} \vec{F} \cdot d\vec{A} = \int_{S_1} \vec{G} \cdot d\vec{S}.$$

Similarly, if we define the t and u components of $\vec{G}$ by

$$G_2 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial s} \quad \text{and} \quad G_3 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t},$$

then

$$\int_{A_i} \vec{F} \cdot d\vec{A} = \int_{S_i} \vec{G} \cdot d\vec{S}, \quad i = 2, \ldots, 6.$$

(See Problem 4.) Adding the integrals for all the faces, we find that

$$\int_A \vec{F} \cdot d\vec{A} = \int_S \vec{G} \cdot d\vec{S}.$$

Since we have already proved the Divergence Theorem for the rectangular region W, we have

$$\int_S \vec{G}\cdot d\vec{S} = \int_W \operatorname{div}\vec{G}\; dW,$$

where

$$\operatorname{div}\vec{G} = \frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u}.$$

Problems 5 and 6 on page 66 show that

$$\frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u} = \left|\frac{\partial(x,y,z)}{\partial(s,t,u)}\right| \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\right).$$

So, by the three-variable change of variables formula on page **??**,

$$\begin{aligned}
\int_V \operatorname{div}\vec{F}\; dV &= \int_V \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\right) dx\,dy\,dz \\
&= \int_W \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\right) \left|\frac{\partial(x,y,z)}{\partial(s,t,u)}\right| ds\,dt\,du \\
&= \int_W \left(\frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u}\right) ds\,dt\,du \\
&= \int_W \operatorname{div}\vec{G}\; dW.
\end{aligned}$$

In summary, we have shown that

$$\int_A \vec{F}\cdot d\vec{A} = \int_S \vec{G}\cdot d\vec{S}$$

and

$$\int_V \operatorname{div}\vec{F}\; dV = \int_W \operatorname{div}\vec{G}\; dW.$$

By the Divergence Theorem for rectangular solids, the right hand sides of these equations are equal, so the left hand sides are equal also. This proves the Divergence Theorem for the curved region V.

Pasting Regions Together

As in the proof of Green's Theorem, we prove the Divergence Theorem for more general regions by pasting smaller regions together along common faces. Suppose the solid region V is formed by pasting together solids V_1 and V_2 along a common face, as in Figure M.53.

The surface A which bounds V is formed by joining the surfaces A_1 and A_2 which bound V_1 and V_2, and then deleting the common face. The outward flux integral of a vector field $\vec{F}$ through

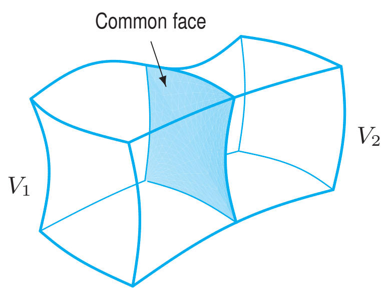

Figure M.53: Region V formed by pasting together V_1 and V_2

A_1 includes the integral across the common face, and the outward flux integral of $\vec{F}$ through A_2 includes the integral over the same face, but oriented in the opposite direction. Thus, when we add the integrals together, the contributions from the common face cancel, and we get the flux integral through A. Thus we have

$$\int_A \vec{F} \cdot d\vec{A} = \int_{A_1} \vec{F} \cdot d\vec{A} + \int_{A_2} \vec{F} \cdot d\vec{A}.$$

But we also have

$$\int_V \operatorname{div} \vec{F}\, dV = \int_{V_1} \operatorname{div} \vec{F}\, dV + \int_{V_2} \operatorname{div} \vec{F}\, dV.$$

So the Divergence Theorem for V follows from the Divergence Theorem for V_1 and V_2. Hence we have proved the Divergence Theorem for any region formed by pasting together regions that can be smoothly parameterized by rectangular solids.

Example 1 Let V be a spherical ball of radius 2, centered at the origin, with a concentric ball of radius 1 removed. Using spherical coordinates, show that the proof of the Divergence Theorem we have given applies to V.

Solution We cut V into two hollowed hemispheres like the one shown in Figure M.54, W. In spherical coordinates, W is the rectangle $1 \le \rho \le 2$, $0 \le \phi \le \pi$, $0 \le \theta \le \pi$. Each face of this rectangle becomes part of the boundary of W. The faces $\rho = 1$ and $\rho = 2$ become the inner and outer hemispherical surfaces that form part of the boundary of W. The faces $\theta = 0$ and $\theta = \pi$ become the two halves of the flat part of the boundary of W. The faces $\phi = 0$ and $\phi = \pi$ become line segments along the z-axis. We can form V by pasting together two solid regions like W along the flat surfaces where $\theta =$ constant.

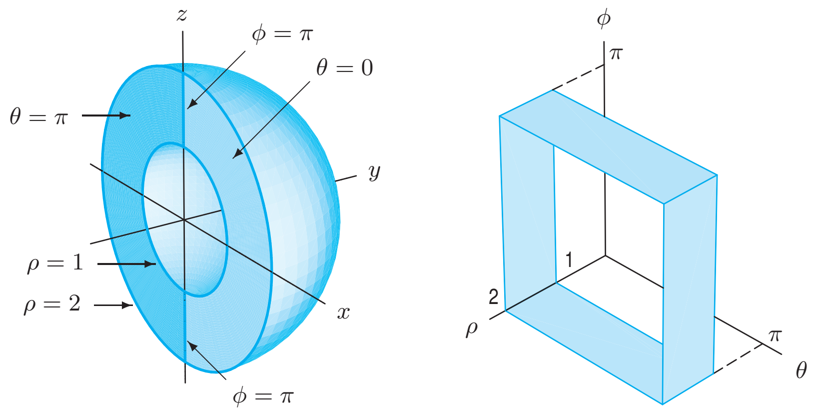

Figure M.54: The hollow hemisphere W and the corresponding rectangular region in $\rho\theta\phi$-space

Proof of Stokes' Theorem

Consider an oriented surface A, bounded by the curve B. We want to prove Stokes' Theorem:

$$\int_A \operatorname{curl} \vec{F} \cdot d\vec{A} = \int_B \vec{F} \cdot d\vec{r}.$$

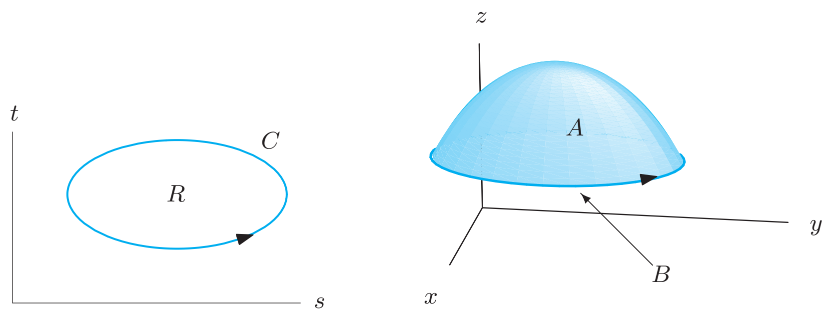

Figure M.55: A region R in the st-plane and the corresponding surface A in xyz-space; the curve C corresponds to the boundary of B

We suppose that A has a smooth parameterization $\vec{r} = \vec{r}(s,t)$, so that A corresponds to a region R in the st-plane, and B corresponds to the boundary C of R. See Figure M.55. We prove Stokes' Theorem for the surface A and a vector field $\vec{F}$ by expressing the integrals on both sides of the theorem in terms of s and t, and using Green's Theorem in the st-plane.

First, we convert the line integral $\int_B \vec{F} \cdot d\vec{r}$ into a line integral around C:

$$\int_B \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \frac{\partial \vec{r}}{\partial s}\, ds + \vec{F} \cdot \frac{\partial \vec{r}}{\partial t}\, dt.$$

So if we define a 2-dimensional vector field $\vec{G} = (G_1, G_2)$ on the st-plane by

$$G_1 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \quad \text{and} \quad G_2 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial t},$$

then

$$\int_B \vec{F} \cdot d\vec{r} = \int_C \vec{G} \cdot d\vec{s},$$

using $\vec{s}$ to denote the position vector of a point in the st-plane.

What about the flux integral $\int_A \operatorname{curl} \vec{F} \cdot d\vec{A}$ that occurs on the other side of Stokes' Theorem? In terms of the parameterization,

$$\int_A \operatorname{curl} \vec{F} \cdot d\vec{A} = \int_R \operatorname{curl} \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t}\, ds\, dt.$$

In Problem 7 on page 67 we show that

$$\operatorname{curl} \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t} = \frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t}.$$

Hence

$$\int_A \operatorname{curl} \vec{F} \cdot d\vec{A} = \int_R \left(\frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t} \right) ds\, dt.$$

We have already seen that

$$\int_B \vec{F} \cdot d\vec{r} = \int_C \vec{G} \cdot d\vec{s}.$$

By Green's Theorem, the right-hand sides of the last two equations are equal. Hence the left-hand sides are equal as well, which is what we had to prove for Stokes' Theorem.

Problems for Section M

1. Let W be a solid circular cylinder along the z-axis, with a smaller concentric cylinder removed. Parameterize W by a rectangular solid in $r\theta z$-space, where r, θ, and z are cylindrical coordinates.

2. In this section we proved the Divergence Theorem using the coordinate definition of divergence. Now we use the Divergence Theorem to show that the coordinate definition is the same as the geometric definition. Suppose $\vec{F}$ is smooth in a neighborhood of (x_0, y_0, z_0), and let U_R be the ball of radius R with center (x_0, y_0, z_0). Let m_R be the minimum value of $\operatorname{div} \vec{F}$ on U_R and let M_R be the maximum value.

(a) Let S_R be the sphere bounding U_R. Show that

$$m_R \le \frac{\int_{S_R} \vec{F} \cdot d\vec{A}}{\text{Volume of } U_R} \le M_R.$$

(b) Explain why we can conclude that

$$\lim_{R\to 0} \frac{\int_{S_R} \vec{F} \cdot d\vec{A}}{\text{Volume of } U_R} = \operatorname{div} \vec{F}\,(x_0, y_0, z_0).$$

(c) Explain why the statement in part (b) remains true if we replace U_R with a cube of side R, centered at (x_0, y_0, z_0).

Problems 3–6 fill in the details of the proof of the Divergence Theorem.

3. Figure M.52 on page 61 shows the solid region V in xyz-space parameterized by a rectangular solid W in stu-space using the change of coordinates

$$\vec{r} = \vec{r}\,(s, t, u), \quad a \le s \le b, c \le t \le d, e \le u \le f.$$

Suppose that $\dfrac{\partial \vec{r}}{\partial s} \cdot \left(\dfrac{\partial \vec{r}}{\partial t} \times \dfrac{\partial \vec{r}}{\partial u} \right)$ is positive.

(a) Let A_1 be the face of V corresponding to the face $s = a$ of W. Show that $\dfrac{\partial \vec{r}}{\partial s}$, if it is not zero, points into W.

(b) Show that $-\dfrac{\partial \vec{r}}{\partial t} \times \dfrac{\partial \vec{r}}{\partial u}$ is an outward pointing normal on A_1.

(c) Find an outward pointing normal on A_2, the face of V where $s = b$.

4. Show that for the other five faces of the solid V in the proof of the Divergence Theorem (see page 62):

$$\int_{A_i} \vec{F} \cdot d\vec{A} = \int_{S_i} \vec{G} \cdot d\vec{S}, \quad i = 2, 3, 4, 5, 6.$$

5. Suppose that $\vec{F}$ is a vector field and that $\vec{a}$, $\vec{b}$, and $\vec{c}$ are vectors. In this problem we prove the formula

$$\operatorname{grad}(\vec{F} \cdot \vec{b} \times \vec{c}\,) \cdot \vec{a} + \operatorname{grad}(\vec{F} \cdot \vec{c} \times \vec{a}\,) \cdot \vec{b} \\ + \operatorname{grad}(\vec{F} \cdot \vec{a} \times \vec{b}\,) \cdot \vec{c} = (\vec{a} \cdot \vec{b} \times \vec{c}\,) \operatorname{div} \vec{F}.$$

(a) Interpretating the divergence as flux density, explain why the formula makes sense. [Hint: Consider the flux out of a small parallelepiped with edges parallel to $\vec{a}$, $\vec{b}$, $\vec{c}$.]

(b) Say how many terms there are in the expansion of the left hand side of the formula in Cartesian coordinates, without actually doing the expansion.

(c) Write down all the terms on the left hand side that contain $\partial F_1/\partial x$. Show that these terms add up to $\vec{a} \cdot \vec{b} \times \vec{c}\, \dfrac{\partial F_1}{\partial x}$.

(d) Write down all the terms that contain $\partial F_1/\partial y$. Show that these add to zero.

(e) Explain how the expressions involving the other seven partial derivatives will work out, and how this verifies that the formula holds.

6. Let $\vec{F}$ be a smooth vector field in 3-space, and let

$$x = x(s, t, u), \quad y = y(s, t, u), \quad z = z(s, t, u)$$

be a smooth change of variables, which we will write in vector form as

$$\vec{r} = \vec{r}\,(s, t, u) = x(s, t, u)\vec{i} + y(s, t, u)\vec{j} + z(s, t, u)\vec{k}.$$

Define a vector field $\vec{G} = (G_1, G_2, G_3)$ on stu-space by

$$G_1 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial t} \times \frac{\partial \vec{r}}{\partial u} \qquad G_2 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial s}$$
$$G_3 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t}.$$

(a) Show that

$$\frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u} = \frac{\partial \vec{F}}{\partial s} \cdot \frac{\partial \vec{r}}{\partial t} \times \frac{\partial \vec{r}}{\partial u} \\ + \frac{\partial \vec{F}}{\partial t} \cdot \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial s} + \frac{\partial \vec{F}}{\partial u} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t}.$$

(b) Let $\vec{r}_0 = \vec{r}\,(s_0, t_0, u_0)$, and let

$$\vec{a} = \frac{\partial \vec{r}}{\partial s}(\vec{r}_0), \quad \vec{b} = \frac{\partial \vec{r}}{\partial t}(\vec{r}_0), \quad \vec{c} = \frac{\partial \vec{r}}{\partial u}(\vec{r}_0).$$

Use the chain rule to show that

$$\left(\frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u} \right) \Bigg|_{\vec{r} = \vec{r}_0} = \\ \operatorname{grad}(\vec{F} \cdot \vec{b} \times \vec{c}\,) \cdot \vec{a} + \operatorname{grad}(\vec{F} \cdot \vec{c} \times \vec{a}\,) \cdot \vec{b} \\ + \operatorname{grad}(\vec{F} \cdot \vec{a} \times \vec{b}\,) \cdot \vec{c}.$$

(c) Use Problem 5 to show that

$$\frac{\partial G_1}{\partial s} + \frac{\partial G_2}{\partial t} + \frac{\partial G_3}{\partial u} = \\ \left| \frac{\partial(x, y, z)}{\partial(s, t, u)} \right| \left(\frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} \right).$$

7. This problem completes the proof of Stokes' Theorem. Let $\vec{F}$ be a smooth vector field in 3-space, and let S be a surface parameterized by $\vec{r} = \vec{r}(s,t)$. Let $\vec{r}_0 = \vec{r}(s_0,t_0)$ be a fixed point on S. We define a vector field in st-space as on page 65:

$$G_1 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \qquad G_2 = \vec{F} \cdot \frac{\partial \vec{r}}{\partial t}.$$

(a) Let $\vec{a} = \frac{\partial \vec{r}}{\partial s}(\vec{r}_0)$, $\quad \vec{b} = \frac{\partial \vec{r}}{\partial t}(\vec{r}_0)$. Show that

$$\frac{\partial G_1}{\partial t}(\vec{r}_0) - \frac{\partial G_2}{\partial s}(\vec{r}_0) =$$

$$\operatorname{grad}(\vec{F} \cdot \vec{a}) \cdot \vec{b} - \operatorname{grad}(\vec{F} \cdot \vec{b}) \cdot \vec{a}.$$

(b) Use Problem **??** on page **??** of the textbook to show

$$\operatorname{curl} \vec{F} \cdot \frac{\partial \vec{r}}{\partial s} \times \frac{\partial \vec{r}}{\partial t} = \frac{\partial G_2}{\partial s} - \frac{\partial G_1}{\partial t}.$$