Math 111
Calculus I

R. Mayer

$\cdots$ there was far more imagination in the head of Archimedes than in that of Homer.

Voltaire[46, page 170]

Contents

Acknowledgements

I would like to thank Joe Buhler, David Perkinson, Jamie Pommersheim, Rao Potluri, Joe Roberts, Jerry Shurman, and Steve Swanson for suggesting numerous improvements to earlier drafts of these notes. I would also like to thank Cathy D'Ambrosia for converting my handwritten notes into LaTeX.

Chapter 0

Introduction

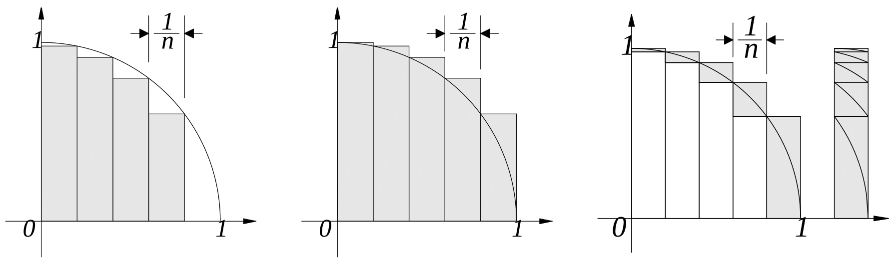
An Overview of the Course

In the first part of these notes we consider the problem of calculating the areas of various plane figures. The technique we use for finding the area of a figure A will be to construct a sequence I_n of sets contained in A, and a sequence O_n of sets containing A, such that

1. The areas of I_n and O_n are easy to calculate.

2. When n is large then both I_n and O_n are in some sense "good approximations" for A.

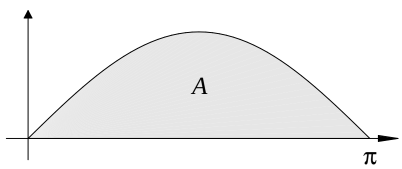
Then by examining the areas of I_n and O_n we will determine the area of A. The figure below shows the sorts of sets we might take for I_n and O_n in the case where A is the set of points in the first quadrant inside of the circle $x^2+y^2 = 1$.



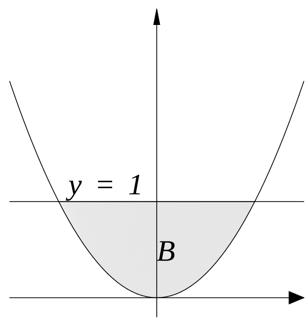
I_n $\qquad$ O_n $\qquad$ $\text{area}(O_n) - \text{area}(I_n) = \frac{1}{n}$

In this example, both of the sets I_n and O_n are composed of a finite number of rectangles of width $\frac{1}{n}$, and from the equation of the circle we can calculate the heights of the rectangles, and hence we can find the areas of I_n and O_n. From the third figure we see that area$(O_n)-$ area$(I_n) = \frac{1}{n}$. Hence if $n = 100000$, then either of the numbers area(I_n) or area(O_n) will give the area of the quarter-circle with an error of no more than 10^{-5}. This calculation will involve taking many square roots, so you probably would not want to carry it out by hand, but with the help of a computer you could easily find the area of the circle to five decimals accuracy. However no amount of computing power would allow you to get thirty decimals of accuracy from this method in a lifetime, and we will need to develop some theory to get better approximations.

In some cases we can find exact areas. For example, we will show that the area of one arch of a sine curve is 2, and the area bounded by the parabola $y = x^2$ and the line $y = 1$ is $\frac{4}{3}$.



$y = \sin(x)$ area$(A) = 2$



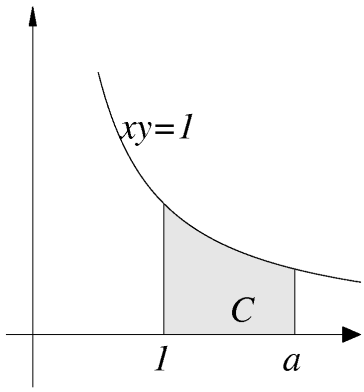
$y = x^2$ area$(B) = \frac{4}{3}$

However in other cases the areas are not simply expressible in terms of known numbers. In these cases we define certain numbers in terms of areas, for example we will define

$$\pi = \text{the area of a circle of radius 1,}$$

and for all numbers $a > 1$ we will define

$$\begin{aligned}\ln(a) &= \text{the area of the region bounded by the curves}\\ &\quad y = 0,\ xy = 1,\ x = 1,\ \text{ and } x = a.\end{aligned}$$



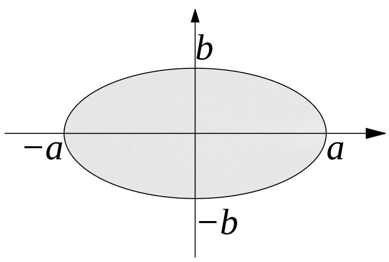
area(C) = ln(a)

We will describe methods for calculating these numbers to any degree of accuracy, and then we will consider them to be known numbers, just as you probably now think of $\sqrt{2}$ as being a known number. (Many calculators calculate these numbers almost as easily as they calculate square roots.) The numbers $\ln(a)$ have many interesting properties which we will discuss, and they have many applications to mathematics and science.

Often we consider general classes of figures, in which case we want to find a simple formula giving areas for all of the figures in the class. For example we will express the area of the ellipse bounded by the curve whose equation is

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

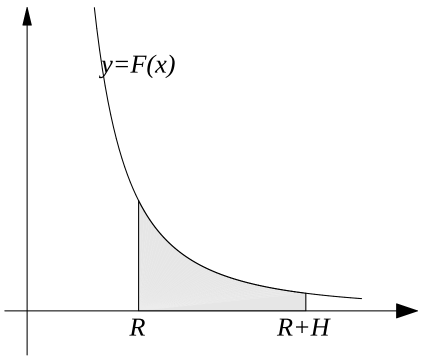
by means of a simple formula involving a and b.



The mathematical tools that we develop for calculating areas, (i.e. the theory of *integration*) have many applications that seem to have little to do with area. Consider a moving object that is acted upon by a known force $F(x)$ that depends on the position x of the object. (For example, a rocket propelled upward from the surface of the moon is acted upon by the moon's gravitational attraction, which is given by

$$F(x) = \frac{C}{x^2},$$

where x is the distance from the rocket to the center of the moon, and C is some constant that can be calculated in terms of the mass of the rocket and known information.) Then the amount of work needed to move the object from a position $x = x_0$ to a position $x = x_1$ is equal to the area of the region bounded by the lines $x = x_0$, $x = x_1$, $y = 0$ and $y = F(x)$.



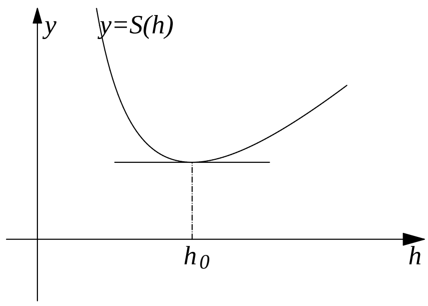
Work is represened by an area

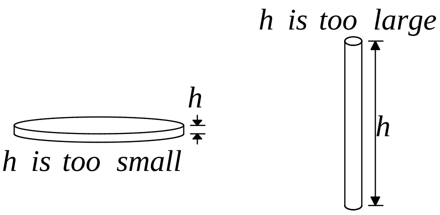
In the case of the moon rocket, the work needed to raise the rocket a height H above the surface of the moon is the area bounded by the lines $x = R$, $x = R + H$, $y = 0$, and $y = \frac{C}{x^2}$, where R is the radius of the moon. After we have developed a little bit of machinery, this will be an easy area to calculate. The amount of work here determines the amount of fuel necessary to raise the rocket.

Some of the ideas used in the theory of integration are thousands of years old. Quite a few of the technical results in the calculations presented in these notes can be found in the writings of Archimedes(287–212 B.C.), although the way the ideas are presented here is not at all like the way they are presented by Archimedes.

In the second part of the notes we study the idea of *rate of change*. The ideas used in this section began to become common in early seventeenth century, and they have no counterpart in Greek mathematics or physics. The problems considered involve describing motions of moving objects (e.g. cannon balls or planets), or finding tangents to curves. An important example of a rate of change is *velocity*. The problem of what is meant by the velocity of a moving object at a given instant is a delicate one. At a particular instant of time, the object occupies just one position in space. Hence during that instant the object does not move. If it does not move, it is at rest. If it is at rest, then its velocity must be 0(?)

The ability to find tangents to curves allows us to find maximum and minimum values of functions. Suppose I want to design a tin can that holds 1000 cc., and requires a minimum amount of tin. It is not hard to find a function S such that for each positive number h, the total surface area of a can with height h and volume 1000 is equal to $S(h)$. The graph of S has the general shape shown in the figure, and the minimum surface area corresponds to the height h_0 shown in the figure. This value h_0 corresponds to the point on the graph of S where the tangent line is horizontal, i.e. where the slope of the tangent is zero. From the formula for $S(h)$ we will be able to find a formula for the slope of the tangent to the graph of S at an arbitrary height h, and to determine when the slope is zero. Thus we will find h_0.





Badly designed cans

The tool for solving rate problems is the *derivative*, and the process of calculating derivatives is called *differentiation*. (There are two systems of notation working here. The term *differential* was introduced by Gottfried Leibniz(1646–1716) to describe a concept that later developed into what Joseph Louis Lagrange(1736–1813) called the *derived function*. From Lagrange we get our word *derivative*, but the older name due to Leibniz is still used to describe the general theory – from which differentials in the sense of Leibniz have been banished.) The idea of derivative (or fluxion or differential) appears in the work of Isaac Newton(1642–1727) and of Leibniz, but can be found in various disguises in the work of a number of earlier mathematicians.

As a rule, it is quite easy to calculate the velocity and acceleration of a moving object, if a formula for the position of the object at an arbitrary time is known. However usually no such formula is obvious. Newton's Second Law states that the acceleration of a moving object is proportional to the sum of the forces acting on the object, divided by the mass of the object. Now often we have a good idea of what the forces acting on an object are, so we know the acceleration. The interesting problems involve calculating velocity and position from acceleration. This is a harder problem than the problem going in the opposite direction, but we will find ways of solving this problem in many cases. The natural statements of many physical laws require the notion of derivative for their statements. According to Salomon Bochner

> The mathematical concept of derivative is a master concept, one of the most creative concepts in analysis and also in human cognition altogether. Without it there would be no velocity or acceleration or momentum, no density of mass or electric charge or any other density, no gradient of a potential and hence no concept of potential in any part of physics, no wave equation; no mechanics no physics, no technology, nothing[11, page 276].

At the time that ideas associated with differentiation were being developed, it was widely recognized that a logical justification for the subject was completely lacking. However it was generally agreed that the results of the calculations based on differentiation were correct. It took more than a century before a logical basis for derivatives was developed, and the concepts of *function* and *real number* and *limit* and *continuity* had to be developed before the foundations could be described. The story is probably not complete. The modern "constructions" of real numbers based on a general theory of "sets" appear to me to be very vague, and more closely related to philosophy than to mathematics. However in these notes we will not worry about the foundations of the real numbers. We will assume that they are there waiting for us to use, but we will need to discuss the concepts of function, limit and continuity in order to get our results.

The *fundamental theorem of the calculus* says that the theory of integration, and the theory of differentiation are very closely related, and that differentiation techniques can be used for solving integration problems, and vice versa. The fundamental theorem is usually credited to Newton and Leibniz independently, but it can be found in various degrees of generality in a number of earlier writers. It was an idea floating in the air, waiting to be discovered at the close of the seventeenth century.

Prerequisites

The prerequisites for this course are listed in appendix C. You should look over this appendix, and make sure that everything in it is more or less familiar to you. If you are unfamiliar with much of this material, you might want to discuss with your instructor whether you are prepared to take the course. It will be helpful to have studied some trigonometry, but all of the trigonometry used in these notes will be developed as it is needed.

You should read these notes carefully and critically. There are quite a few cases where I have tried to trick you by giving proofs that use unjustified assumptions. In these cases I point out that there is an error after the proof is complete, and either give a new proof, or add some hypotheses to the statement of the theorem. If there is something in a proof that you do not understand, there is a good chance that the proof is wrong.

Exercises and Entertainments

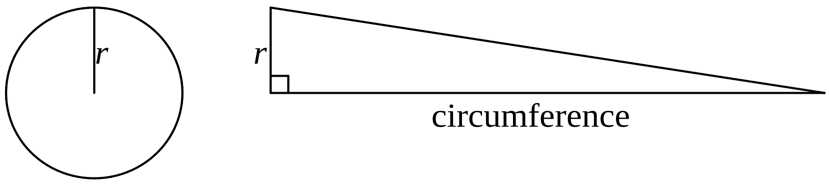
The exercises are an important part of the course. Do not expect to be able to do all of them the first time you try them, but you should understand them after they have been discussed in class. Some important theorems will be proved in the exercises. There are hints for some of the questions in appendix A, but you should not look for a hint unless you have made some effort to answer a question.

Sections whose titles are marked by an asterisk (e.g. section 2.6) are not used later in the notes, and may be omitted. Hovever they contain really neat material, so you will not want to omit them.

In addition to the exercises, there are some questions and statements with the label "entertainment". These are for people who find them entertaining. They require more time and thought than the exercises. Some of them are more metaphysical than mathematical, and some of them require the use of a computer or a programmable calculator. If you do not find the entertainments entertaining, you may ignore them. Here is one to start you off.

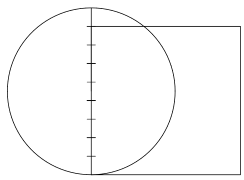
1 Entertainment (Calculation of π.) . The area of a circle of radius 1 is denoted by π. Calculate π as accurately as you can.

Archimedes showed that π is half of the circumference of a circle of radius 1. More precisely, he showed that the area of a circle is equal to the area of a triangle whose base is equal to the circumference of the circle, and whose altitude is equal to the radius of the circle. If we take a circle of radius 1, we get the result stated.



You should assume Archimedes' theorem, and then entertainment 1 is equivalent to the problem of calculating the circumference of a circle as accurately as you can. An answer to this problem will be a pair of rational numbers b and c, together with an argument that $b < \pi$ and $\pi < c$. It is desired to make the difference $c - b$ as small as possible.

This problem is very old. The Rhind Papyrus[16, page 92] (c. 1800 B.C.?) contains the following rule for finding the area of a circle:



RULE I: Divide the diameter of the circle into nine equal parts, and form a square whose side is equal to eight of the parts. Then the area of the square is equal to the area of the circle.

The early Babylonians (1800-1600BC) [38, pages 47 and 51] gave the following rule:

RULE II: The area of a circle is 5/60th of the square of the circumference of the circle.

Archimedes (287–212 B.C.) proved that the circumference of a circle is three times the diameter plus a part smaller than one seventh of the diameter, but greater than 10/71 of the diameter[3, page 134]. In fact, by using only elementary geometry, he gave a method by which π can be calculated to any degree of accuracy by someone who can calculate square roots to any degree of accuracy. We do not know how Archimedes calculated square roots, but people have tried to figure out what method he used by the form of his approximations. For example he says with no justification that

$$\frac{265}{153} < \sqrt{3} < \frac{1351}{780}$$

and

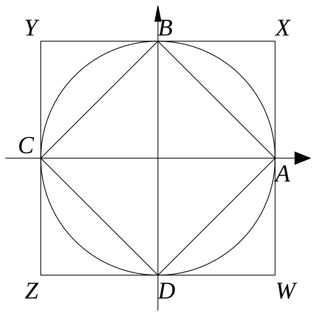
$$\sqrt{3380929} < 1838\frac{9}{11}.$$

By using your calculator you can easily verify that these results are correct. Presumably when you calculate π you will use a calculator or computer to estimate any square roots you need. This immediately suggests a new problem.

2 Entertainment (Square root problem.) Write, or at least describe, a computer program that will calculate square roots to a good deal of accuracy. This program should use only the standard arithmetic operations and the constructions available in all computer languages, and should not use any special functions like square roots or logarithms. An answer to this question must include some sort of explanation of why the method works.

Zǔ Chōngzhī (429–500 A.D.) stated that π is between 3.1415926 and 3.1415927, and gave 355/113 as a good approximation to π.[47, page 82]

Here is a first approximation to π. Consider a circle of radius 1 with center at $(0,0)$, and inscribe inside of it a square $ABCD$ of side s with vertices at $(1,0),(0,1),(-1,0)$ and $(0,-1)$. Then by the Pythagorean theorem, $s^2 = 1^2 + 1^2 = 2$. But s^2 is the area of the square $ABCD$, and since $ABCD$ is contained inside of the circle we have



$$2 = \text{Area of inscribed square} < \text{Area of circle} = \pi.$$

Consider also the circumscribed square $WXYZ$ with horizontal and vertical sides. This square has side 2, and hence has area 4. Thus, since the circle is contained in $WXYZ$,

$$\pi = \text{area of circle} < \text{area}(WXYZ) = 4.$$

It now follows that $2 < \pi < 4$.

A number of extraordinary formulas for π are given in a recent paper on *How to Compute One Billion Digits of Pi*[12]. One amazing formula given in this paper is the following result

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n)!}{(n!)^4} \frac{[1103 + 26390n]}{396^{4n}},$$

which is due to S. Ramanujan(1887–1920)[12, p 201,p 215]. The reciprocal of the zeroth term of this sum i.e.

$$\frac{9801}{1103\sqrt{8}}$$

gives a good approximation to π (see exercise 4).

3 Exercise. The formulas described in RULES I and II above each determine an approximate value for π. Determine the two approximate values. Explain your reasoning.

4 Exercise. Use a calculator to find the value of

$$\frac{9801}{1103\sqrt{8}},$$

and compare this with the correct value of π, which is $3.14159265358979\ldots$.

Chapter 1

Some Notation for Sets

A *set* is any collection of *objects.* Usually the objects we consider are things like numbers, points in the plane, geometrical figures, or functions. Sets are often described by listing the objects they contain inside curly braces, for example

$$\begin{aligned} A &= \{1, 2, 3, 4\}, \\ B &= \{2, 3, 4\}, \\ C &= \{4, 3, 3, 2\}, \\ D &= \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}\}. \end{aligned}$$

There are a few sets that occur very often in mathematics, and that have special names:

$$\begin{aligned} \mathbf{N} &= \text{the set of all natural numbers} = \{0, 1, 2, 3, \ldots\}. \\ \mathbf{Z} &= \text{the set of all integers} = \{0, -1, 1, -2, 2, \ldots\}. \\ \mathbf{Z}^+ &= \text{the set of all positive integers.} \\ \mathbf{Z}^- &= \text{the set of all negative integers.} \\ \mathbf{R} &= \text{the set of all real numbers.} \\ \mathbf{R}^+ &= \text{the set of all positive real numbers.} \\ \mathbf{R}^- &= \text{the set of all negative real numbers.} \\ \mathbf{R}^2 &= \text{the set of all points in the plane.} \\ \mathbf{Q} &= \text{the set of all rational numbers.} \\ \mathbf{Q}^+ &= \text{the set of all positive rational numbers.} \\ \mathbf{Q}^- &= \text{the set of all negative rational numbers.} \\ \emptyset &= \text{the empty set} = \text{the set containing no elements.} \end{aligned}$$

A rational number is a number that can be expressed as a quotient of two integers. Thus a real number x is rational if and only if there exist integers a and b with $b \neq 0$ such that $x = a/b$.

The terms "point in the plane" and "ordered pair of real numbers" are taken to be synonymous. I assume that you are familiar with the usual representation of points in the plane by pairs of numbers, and the usual way of representing geometrical objects by equations and inequalities.

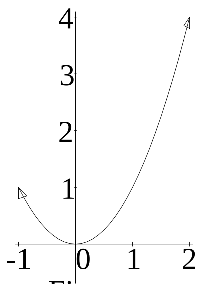


Figure a

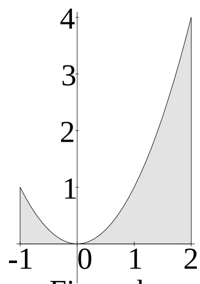


Figure b

Thus the set of points (x, y) such that $y = x^2$ is represented in figure a, and the set of points (x, y) such that $-1 \leq x \leq 2$ and $0 \leq y \leq x^2$ is represented in figure b. The arrowheads in figure a indicate that only part of the figure has been drawn.

The objects in a set S are called *elements of* S or *points in* S. If x is an object and S is a set then

$$x \in S \text{ means that } x \text{ is an element of } S,$$

and

$$x \notin S \text{ means that } x \text{ is not an element of } S.$$

Thus in the examples above

$2 \in A$.	$2/6 \in D$.	$1 \notin C$.	$0 \notin \mathbf{Z}^+$.
$0 \notin \emptyset$.	$\emptyset \notin A$.	$6/3 \in \mathbf{N}$.	$0 \in \mathbf{N}$.

To see that $\emptyset \notin A$, observe that A has exactly four elements, and none of these elements is $\emptyset$.

Let S and T be sets. We say that S is a *subset* of T and write $S \subset T$ if and only if every element in S is also in T. Two sets are considered to be equal if and only if they have exactly the same elements. Thus

$$S = T \text{ means } (S \subset T \text{ and } T \subset S).$$

You can show that two sets are *not* equal, by finding an element in one of the sets that is not in the other.

In the examples above, $B \subset A$ and $B = C$. For every set S we have

$$S \subset S \text{ and } \emptyset \subset S.$$

Also

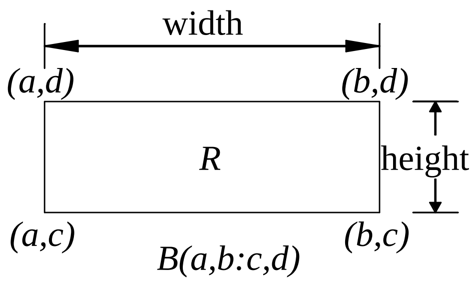
$$\begin{array}{cc} 1 \in \{1,2,3\}. & \emptyset \notin \{1,2,3\}. \\ \{1\} \subset \{1\}. & \emptyset \subset \{1,2,3\}. \\ \{1\} \notin \{1\}. & 1 \in \{1\}. \end{array}$$

The idea of set was introduced into mathematics by Georg Cantor near the end of the nineteenth century. Since then it has become one of the most important ideas in mathematics. In these notes we use very little from the *theory* of sets, but the *language* of sets will be very evident.

1.1 Definition (Box, width, height, area.) Let a, b, c, d be real numbers with $a \leq b$ and $c \leq d$. We define the set $B(a, b\colon c, d)$ by

$$\begin{aligned} B(a, b\colon c, d) \;\; = \;\; & \text{the set of points } (x, y) \text{ in } \mathbf{R}^2 \text{ such that} \\ & a \leq x \leq b \;\text{ and } c \leq y \leq d. \end{aligned}$$

A set of this type will be called a *box*. If $R = B(a, b : c, d)$, then we will refer to the number $b - a$ as the *width of* R, and we refer to $d - c$ as the *height of* R.



The *area* of the box $B(a, b : c, d)$ is the number

$$\text{area}(B(a, b : c, d)) = (b - a)(d - c).$$

Remark: Notice that in the definition of box, the inequalities are "$\leq$" and not "$<$". The choice of which sort of inequality to use is somewhat arbitrary, but some of the assertions we will be making about boxes would turn out to be false if the boxes did not contain their boundaries.

In Euclid's geometry no distinction is made between sets that contain their boundaries and sets that do not. In fact the early Greek geometers did not think in terms of sets at all. Aristotle maintained that

> A line cannot be made up of points, seeing that a line is a continuous thing, and a point is indivisible[25, page 123].

The notion that geometric figures are sets of points is a very modern one. Also the idea that area is a *number* has no counterpart in Euclid's geometry, and in fact Euclid does not talk about area at all. He makes statements like

> Triangles which are on equal bases and in the same parallels are equal to one another[17, vol I page 333].

We interpret "are equal to one another" to mean "have equal areas", but Euclid does not define "equal" or mention "area".

1.2 Definition (Unions and Intersections.) Let $F = \{S_1, \cdots, S_n\}$ be a set of sets. The *union* of the sets $S_1, \cdots, S_n$ is defined to be the set of all points x that belong to at least one of the sets $S_1, \cdots, S_n$. This union is denoted by

$$S_1 \cup S_2 \cup \cdots \cup S_n$$

or by

$$\bigcup_{i=1}^{n} S_i. \tag{1.3}$$

The *intersection* of the sets $S_1, S_2, \cdots, S_n$ is defined to be the set of points x that are in every one of the sets S_i. This intersection is denoted by

$$S_1 \cap S_2 \cap \cdots \cap S_n$$

or by

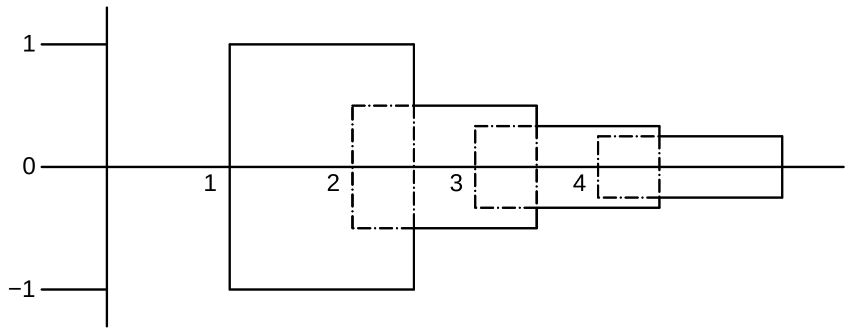
$$\bigcap_{i=1}^{n} S_i. \tag{1.4}$$

The index i in equations 1.3 and 1.4 is called a *dummy index* and it can be replaced by any symbol that does not have a meaning assigned to it. Thus,

$$\bigcup_{i=1}^{n} S_i = \bigcup_{k=1}^{n} S_k = \bigcup_{t=1}^{n} S_t,$$

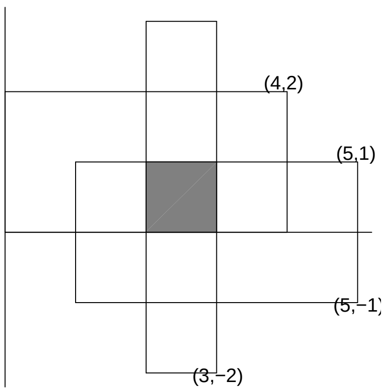
but expressions such as $\bigcup_{n=1}^{n} S_n$ or $\bigcup_{3=1}^{n} S_3$ will be considered to be ungrammatical.

1.5 Example. For $i \in \mathbf{Z}^+$ let $R_i = B(i, i + \frac{3}{2} : -\frac{1}{i}, \frac{1}{i})$. Then $\bigcup_{i=1}^{4} R_i$ is represented in the figure, and $\bigcap_{i=1}^{4} R_i = \emptyset$. Also $R_1 \cap R_2 = B(2, \frac{5}{2} : -\frac{1}{2}, \frac{1}{2})$.



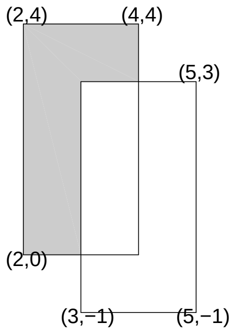
In the figure below,

$$B(0, 4 : 0, 2) \cap B(1, 5 : -1, 1) \cap B(2, 3 : -2, 3) = B(2, 3 : 0, 1).$$



1.6 Definition (Set difference.) If A and B are sets then the *set difference* $A \setminus B$ is the set of all points that are in A but not in B.

In the figure, the shaded region represents $B(2, 4\colon 0, 4) \setminus B(3, 5\colon -1, 3)$.



1.7 Exercise. Explain why it is *not* true that

$$B(2, 4\colon 0, 4) \setminus B(3, 5\colon -1, 3) = B(2, 3\colon 0, 4) \cup B(3, 4\colon 3, 4).$$

I will often use set relations such as

$$A \cup B = (A \setminus B) \cup B$$

or

$$A = (A \setminus B) \cup (A \cap B)$$

without explanation or justification. The second statement says that A consists of the points in A which are not in B together with the points in A that are in B, and I take this and similar statements to be clear.

1.8 Definition (Intervals.) Let a, b be real numbers with $a \leq b$. We define the following subsets of $\mathbf{R}$:

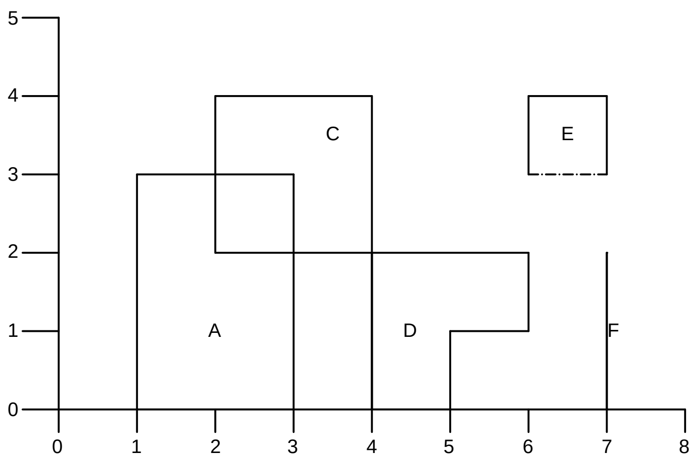
$$\begin{aligned}
(a, b) &= \text{the set of real numbers } x \text{ such that } a < x < b. \\
(a, b] &= \text{the set of real numbers } x \text{ such that } a < x \leq b. \\
[a, b) &= \text{the set of real numbers } x \text{ such that } a \leq x < b. \\
[a, b] &= \text{the set of real numbers } x \text{ such that } a \leq x \leq b. \\
(a, \infty) &= \text{the set of real numbers } x \text{ such that } x > a. \\
[a, \infty) &= \text{the set of real numbers } x \text{ such that } x \geq a. \\
(-\infty, a) &= \text{the set of real numbers } x \text{ such that } x < a. \\
(-\infty, a] &= \text{the set of real numbers } x \text{ such that } x \leq a. \\
(-\infty, \infty) &= \text{the set of real numbers.}
\end{aligned}$$

A subset of $\mathbf{R}$ is called an *interval* if it is equal to a set of one of these nine types. Note that $(a,a) = \emptyset$ and $[a,a] = \{a\}$, so the empty set and a set consisting of just one point are both intervals.

1.9 Definition (End points: open and closed intervals.) If I is a non-empty interval of one of the first four types in the above list, then we will say that the *end points* of I are the numbers a and b. If I is an interval of one of the next four types, then I has the unique end point a. The empty set and the interval $(-\infty,\infty)$ have no end points. An interval is *closed* if it contains all of its end points, and it is *open* if it contains none of its end points.

1.10 Exercise. Let a, b be real numbers with $a < b$. For each of the nine types of interval described in definition 1.8, decide whether an interval of the type is open or closed. (Note that some types are both open and closed, and some types are neither open nor closed.) Is the interval $(0,0]$ open? Is it closed? What about the interval $[0,0]$?

1.11 Exercise. In the figure below, A, C, and F are boxes.
a) Express each of A, C, F in the form $B(?,?:?,?)$.
b) Express D, E, and $A \cap C$ as intersections or unions or set differences of boxes. The dotted edge of E indicates that the edge is missing from the set.
c) Find a box that contains $A \cup C$.

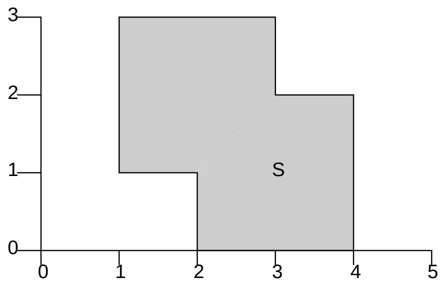


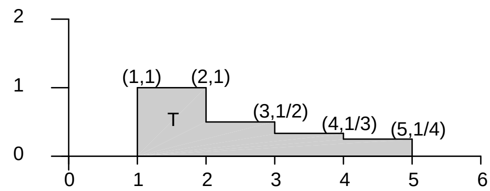
1.12 Exercise. Let S be the set of points (x,y) in $\mathbf{R}^2$ such that $1 \le x \le 4$ and $0 \le y \le \dfrac{1}{x^2}$. Let T be the set of points (x,y) in $\mathbf{R}^2$ such that

$$(x,y) \in B(-1,1:-1,1) \text{ and } xy > 0.$$

Make sketches of the sets S and T.

1.13 Exercise. Describe the sets S and T below in terms of unions or intersections or differences of boxes.



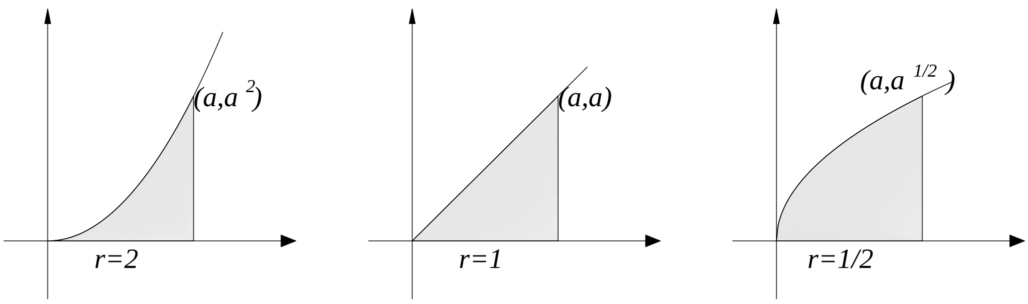


Chapter 2

Some Area Calculations

2.1 The Area Under a Power Function

Let a be a positive number, let r be a positive number, and let S_a^r be the set of points (x, y) in $\mathbf{R}^2$ such that $0 \le x \le a$ and $0 \le y \le x^r$. In this section we will begin an investigation of the area of S_a^r.



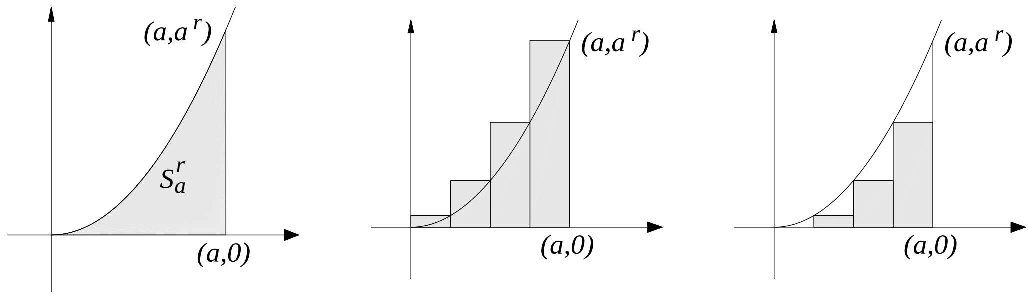
S_a^r for various positive values of r

Our discussion will not apply to negative values of r, since we make frequent use of the fact that for all non-negative numbers x and t

$$(x \le t) \text{ implies that } (x^r \le t^r).$$

Also 0^r is not defined when r is negative.

The figures for the argument given below are for the case $r = 2$, but you should observe that the proof does not depend on the pictures.



$S_a^r \subset \bigcup_{i=1}^4 O_i$ $\qquad$ $\bigcup_{i=1}^4 I_i \subset S_a^r$

Let n be a positive integer, and for $0 \le i \le n$, let $x_i = \dfrac{ia}{n}$.

Then $x_i - x_{i-1} = \dfrac{a}{n}$ for $1 \le i \le n$, so the points x_i divide the interval $[0, a]$ into n equal subintervals. For $1 \le i \le n$, let

$$\begin{aligned} I_i &= B(x_{i-1}, x_i \colon 0, x_{i-1}^r) \\ O_i &= B(x_{i-1}, x_i \colon 0, x_i^r). \end{aligned}$$

If $(x, y) \in S_a^r$, then $x_{i-1} \le x \le x_i$ for some index i, and $0 \le y \le x^r \le x_i^r$, so

$$(x, y) \in B(x_{i-1},\ x_i \colon 0,\ x_i^r) = O_i \text{ for some } i \in \{1, \cdots, n\}.$$

Hence we have

$$S_a^r \subset \bigcup_{i=1}^n O_i,$$

and thus

$$\text{area}(S_a^r) \le \text{area}(\bigcup_{i=1}^n O_i). \tag{2.1}$$

If $(x, y) \in I_i$, then $0 \le x_{i-1} \le x \le x_i \le a$ and $0 \le y \le x_{i-1}^r \le x^r$ so $(x, y) \in S_a^r$. Hence, $I_i \subset S_a^r$ for all i, and hence

$$\bigcup_{i=1}^n I_i \subset S_a^r,$$

so that

$$\text{area}(\bigcup_{i=1}^{n} I_i) \leq \text{area}(S_a^r). \tag{2.2}$$

Now

$$\begin{aligned}\text{area}(I_i) &= \text{area}\left(B(x_{i-1}, x_i\colon 0, x_{i-1}^r)\right) \\ &= (x_i - x_{i-1})x_{i-1}^r = \frac{a}{n}\left(\frac{(i-1)a}{n}\right)^r = \frac{a^{r+1}}{n^{r+1}}(i-1)^r,\end{aligned}$$

and

$$\begin{aligned}\text{area}(O_i) &= \text{area}\left(B(x_{i-1}, x_i\colon 0, x_i^r)\right) \\ &= (x_i - x_{i-1})x_i^r = \frac{a}{n}\left(\frac{ia}{n}\right)^r = \frac{a^{r+1}}{n^{r+1}}i^r.\end{aligned}$$

Since the boxes I_i intersect only along their boundaries, we have

$$\begin{aligned}\text{area}(\bigcup_{i=1}^{n} I_i) &= \text{area}(I_1) + \text{area}(I_2) + \cdots + \text{area}(I_n) \\ &= \frac{a^{r+1}}{n^{r+1}}0^r + \frac{a^{r+1}}{n^{r+1}}1^r + \cdots + \frac{a^{r+1}}{n^{r+1}}(n-1)^r \\ &= \frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + (n-1)^r), \end{aligned} \tag{2.3}$$

and similarly

$$\begin{aligned}\text{area}(\bigcup_{i=1}^{n} O_i) &= \text{area}(O_1) + \text{area}(O_2) + \cdots + \text{area}(O_n) \\ &= \frac{a^{r+1}}{n^{r+1}}1^r + \frac{a^{r+1}}{n^{r+1}}2^r + \cdots + \frac{a^{r+1}}{n^{r+1}}n^r \\ &= \frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + n^r).\end{aligned}$$

Thus it follows from equations (2.1) and (2.2) that

$$\frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + (n-1)^r) \leq \text{area}(S_a^r) \leq \frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + n^r). \tag{2.4}$$

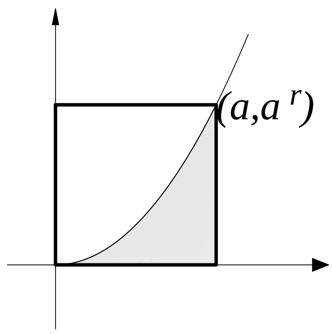
The geometrical question of finding the area of S_a^r has led us to the numerical problem of finding the sum

$$1^r + 2^r + \cdots + n^r.$$

We will study this problem in the next section.

2.5 Definition (Circumscribed box.) Let $\text{cir}(S_a^r)$ be the smallest box containing (S_a^r). i.e.

$$\text{cir}(S_a^r) = B(0, a; 0, a^r) \qquad (r \geq 0).$$



Notice that $\text{area}(\text{cir}(S_a^r)) = a \cdot a^r = a^{r+1}$. Thus equation (2.4) can be written as

$$\frac{(1^r + 2^r + \cdots + (n-1)^r)}{n^{r+1}} \leq \frac{\text{area}(S_a^r)}{\text{area}(\text{cir}(S_a^r))} \leq \frac{(1^r + 2^r + \cdots + n^r)}{n^{r+1}}. \qquad (2.6)$$

Observe that the outside terms in (2.6) do not depend on a.

Now we will specialize to the case when $r = 2$. A direct calculation shows that

$$\begin{aligned} 1^2 &= 1, \\ 1^2 + 2^2 &= 5, \\ 1^2 + 2^2 + 3^2 &= 14, \\ 1^2 + 2^2 + 3^2 + 4^2 &= 30, \\ 1^2 + 2^2 + 3^2 + 4^2 + 5^2 &= 55. \end{aligned} \qquad (2.7)$$

There is a simple (?) formula for $1^2 + 2^2 + \cdots + n^2$, but it is not particularly easy to guess this formula on the basis of these calculations. With the help of my computer, I checked that

$$1^2 + \cdots + 10^2 = 385 \text{ so } \frac{1^2 + \cdots + 10^2}{10^3} = .385$$

$$1^2 + \cdots + 100^2 = 338350 \text{ so } \frac{1^2 + \cdots + 100^2}{100^3} = .33835$$

$$1^2 + \cdots + 1000^2 = 333833500 \text{ so } \frac{1^2 + \cdots + 1000^2}{1000^3} = .3338335$$

Also

$$\begin{aligned} \frac{1^2 + \cdots + 999^2}{1000^3} &= \frac{1^2 + \cdots + 1000^2}{1000^3} - \frac{1000^2}{1000^3} = .3338335 - .001 \\ &= .3328335. \end{aligned}$$

Thus by taking $n = 1000$ in equation (2.6) we see that

$$.332 \leq \frac{\text{area}(S_a^2)}{\text{area}(\text{cir}(S_a^2))} \leq .3339.$$

On the basis of the computer evidence it is very tempting to guess that

$$\text{area}(S_a^2) = \frac{1}{3}\text{area}(\text{cir}(S_a^2)) = \frac{1}{3}a^3.$$

2.2 Some Summation Formulas

We will now develop a formula for the sum

$$1 + 2 + \cdots + n.$$

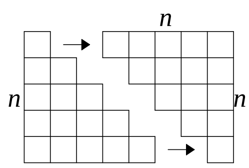


figure a

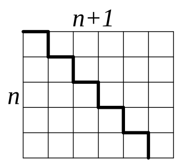


figure b

Figure (a) shows two polygons, each having area $1+2+\cdots+n$. If we slide the two polygons so that they touch, we create a rectangle as in figure (b) whose area is $n(n+1)$. Thus

$$2(1+2+\cdots+n) = n(n+1)$$

i.e.,

$$1+2+\cdots+n = \frac{n(n+1)}{2}. \tag{2.8}$$

The proof just given is quite attractive, and a proof similar to this was probably known to the Pythagoreans in the 6th or 5th centuries B.C. Cf [29, page 30]. The formula itself was known to the Babylonians much earlier than this[45, page 77], but we have no idea how they discovered it.

The idea here is special, and does not generalize to give a formula for $1^2+2^2+\cdots+n^2$. (A nice geometrical proof of the formula for the sum of the first n squares can be found in *Proofs Without Words* by Roger Nelsen[37, page 77], but it is different enough from the one just given that I would not call it a "generalization".) We will now give a second proof of (2.8) that generalizes to give formulas for $1^p+2^p+\cdots+n^p$ for positive integers p. The idea we use was introduced by Blaise Pascal [6, page 197] circa 1654.

For any real number k, we have

$$(k+1)^2 - k^2 = k^2 + 2k + 1 - k^2 = 2k + 1.$$

Hence

$$\begin{aligned}
1^2 - 0^2 &= 2\cdot 0 + 1,\\
2^2 - 1^2 &= 2\cdot 1 + 1,\\
3^2 - 2^2 &= 2\cdot 2 + 1,\\
&\vdots\\
(n+1)^2 - n^2 &= 2\cdot n + 1.
\end{aligned}$$

Add the left sides of these $(n+1)$ equations together, and equate the result to the sum of the right sides:

$$(n+1)^2 - n^2 + \cdots + 3^2 - 2^2 + 2^2 - 1^2 + 1^2 - 0^2 = 2\cdot(1+\cdots+n) + (n+1).$$

In the left side of this equation all of the terms except the first cancel. Thus

$$(n+1)^2 = 2(1+2+\cdots+n) + (n+1)$$

so

$$2(1+2+\cdots+n) = (n+1)^2 - (n+1) = (n+1)(n+1-1) = (n+1)n$$

and

$$1+2+\cdots+n = \frac{n(n+1)}{2}.$$

This completes the second proof of (2.8).

To find $1^2+2^2+\cdots+n^2$ we use the same sort of argument. For any real number k we have

$$(k+1)^3 - k^3 = k^3 + 3k^2 + 3k + 1 - k^3 = 3k^2 + 3k + 1.$$

Hence,

$$\begin{aligned}
1^3 - 0^3 &= 3\cdot 0^2 + 3\cdot 0 + 1,\\
2^3 - 1^3 &= 3\cdot 1^2 + 3\cdot 1 + 1,\\
3^3 - 2^3 &= 3\cdot 2^2 + 3\cdot 2 + 1,\\
&\vdots\\
(n+1)^3 - n^3 &= 3\cdot n^2 + 3\cdot n + 1.
\end{aligned}$$

Next we equate the sum of the left sides of these $n+1$ equations with the sum of the right sides. As before, most of the terms on the left side cancel and we obtain

$$(n+1)^3 = 3(1^2+2^2+\cdots+n^2) + 3(1+2+\cdots+n) + (n+1).$$

We now use the known formula for $1+2+3+\cdots+n$:

$$(n+1)^3 = 3(1^2+2^2+\cdots+n^2) + \frac{3}{2}n(n+1) + (n+1)$$

so

$$\begin{aligned}
3(1^2+2^2+\cdots+n^2) &= (n+1)^3 - \frac{3}{2}n(n+1) - (n+1)\\
&= (n+1)\left((n+1)^2 - \frac{3}{2}n - 1\right)\\
&= (n+1)(n^2+2n+1-\frac{3}{2}n-1)\\
&= (n+1)\left(n^2+\frac{1}{2}n\right) = (n+1)n\left(n+\frac{1}{2}\right)\\
&= \frac{n(n+1)(2n+1)}{2},
\end{aligned}$$

and finally

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2.9}$$

You should check that this formula agrees with the calculations made in (2.7). The argument we just gave can be used to find formulas for $1^3 + 2^3 + \cdots + n^3$, and for sums of higher powers, but it takes a certain amount of stamina to carry out the details. To find $1^3 + 2^3 + \cdots + n^3$, you could begin with

$$(k+1)^4 - k^4 = 4k^3 + 6k^2 + 4k + 1 \text{ for all } k \in \mathbf{R}.$$

Add together the results of this equation for $k = 0, 1, \cdots, n$ and get

$$(n+1)^4 = 4(1^3 + 2^3 + \cdots + n^3) + 6(1^2 + 2^2 + \cdots + n^2) + 4(1 + \cdots + n) + (n+1).$$

Then use equations (2.8) and (2.9) to eliminate $1^2 + 2^2 + \cdots + n^2$ and $1 + \cdots + n$, and solve for $1^3 + 2^3 + \cdots + n^3$.

2.10 Exercise. Complete the argument started above, and find the formula for $1^3 + 2^3 + \cdots + n^3$.

Jacob Bernoulli (1654–1705) considered the general formula for power sums. By using a technique similar to, but slightly different from Pascal's, he constructed the table below. Here $f(1) + f(2) + \cdots f(n)$ is denoted by $\int f(n)$, and $*$ denotes a missing term: Thus the $*$ in the fourth line of the table below indicates that there is no n^2 term, i.e. the coefficient of n^2 is zero.)

> Thus we can step by step reach higher and higher powers and with slight effort form the following table.

Sums of Powers

$$
\begin{array}{llllllll}
\int n & = \tfrac{1}{2}nn & +\tfrac{1}{2}n, & & & & & \\
\int nn & = \tfrac{1}{3}n^3 & +\tfrac{1}{2}nn & +\tfrac{1}{6}n, & & & & \\
\int n^3 & = \tfrac{1}{4}n^4 & +\tfrac{1}{2}n^3 & +\tfrac{1}{4}nn, & & & & \\
\int n^4 & = \tfrac{1}{5}n^5 & +\tfrac{1}{2}n^4 & +\tfrac{1}{3}n^3 & * - \tfrac{1}{30}n, & & & \\
\int n^5 & = \tfrac{1}{6}n^6 & +\tfrac{1}{2}n^5 & +\tfrac{5}{12}n^4 & * - \tfrac{1}{12}nn, & & & \\
\int n^6 & = \tfrac{1}{7}n^7 & +\tfrac{1}{2}n^6 & +\tfrac{1}{2}n^5 & * - \tfrac{1}{6}n^3 & * + \tfrac{1}{42}n, & & \\
\int n^7 & = \tfrac{1}{8}n^8 & +\tfrac{1}{2}n^7 & +\tfrac{7}{12}n^6 & * - \tfrac{7}{24}n^4 & * + \tfrac{1}{12}nn, & & \\
\int n^8 & = \tfrac{1}{9}n^9 & +\tfrac{1}{2}n^8 & +\tfrac{2}{3}n^7 & * - \tfrac{7}{15}n^5 & * + \tfrac{2}{9}n^3 & * - \tfrac{1}{30}n, & \\
\int n^9 & = \tfrac{1}{10}n^{10} & +\tfrac{1}{2}n^9 & +\tfrac{3}{4}n^8 & * - \tfrac{7}{10}n^6 & * + \tfrac{1}{2}n^4 & * - \tfrac{3}{20}nn, & \\
\int n^{10} & = \tfrac{1}{11}n^{11} & +\tfrac{1}{2}n^{10} & +\tfrac{5}{6}n^9 & * - 1n^7 & * + 1n^5 & * - \tfrac{1}{2}n^3 & * + \tfrac{5}{66}n.
\end{array}
$$

> Whoever will examine the series as to their regularity may be able to continue the table[9, pages 317–320]. [1]

He then states a rule for continuing the table. The rule is not quite an explicit formula, rather it tells how to compute the next line easily when the previous lines are known.

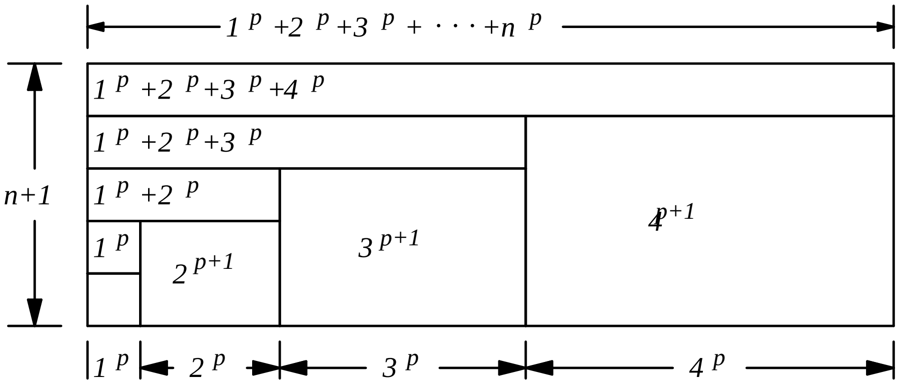
2.11 Entertainment (Bernoulli's problem.) Guess a way to continue the table. Your answer should be explicit enough so that you can actually calculate the next two lines of the table.

A formula for $1^2+2^2+\cdots+n^2$ was proved by Archimedes (287-212 B.C.). (See Archimedes *On Conoids and Spheroids* in [2, pages 107-109]). The formula was known to the Babylonians[45, page 77] much earlier than this in the form

$$1^2+2^2+\cdots+n^2 = (\frac{1}{3}+n\cdot\frac{2}{3})(1+2+\cdots+n).$$

A technique for calculating general power sums has been known since circa 1000 A.D. At about this time Ibn-al-Haitham, gave a method based on the picture below, and used it to calculate the power sums up to $1^4+2^4+\cdots+n^4$. The method is discussed in [6, pages 66–69]

[1] A typographical error in Bernoulli's table has been corrected here.



2.3 The Area Under a Parabola

If S_a^2 is the set of points (x, y) in $\mathbf{R}^2$ such that $0 \le x \le a$ and $0 \le y \le x^2$, then we showed in (2.6) that

$$\frac{1^2 + 2^2 + \cdots + (n-1)^2}{n^3} \le \frac{\text{area}(S_a^2)}{\text{area}(\text{cir}(S_a^2))} \le \frac{1^2 + \cdots + n^2}{n^3}.$$

By (2.9)

$$\begin{aligned} \frac{1^2 + 2^2 + \cdots + n^2}{n^3} &= \frac{n(n+1)(2n+1)}{n^3 \cdot 6} = \frac{1}{3}\left(\frac{n+1}{n}\right)\left(\frac{2n+1}{2n}\right) \\ &= \frac{1}{3}\left(1 + \frac{1}{n}\right)\left(1 + \frac{1}{2n}\right). \end{aligned}$$

Also

$$\begin{aligned} \frac{1^2 + 2^2 + \cdots + (n-1)^2}{n^3} &= \frac{(n-1)n((2(n-1)+1)}{n^3 \cdot 6} = \frac{1}{3}\left(\frac{n-1}{n}\right)\left(\frac{2n-1}{2n}\right) \\ &= \frac{1}{3}\left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{2n}\right), \end{aligned} \tag{2.12}$$

so

$$\frac{1}{3}\left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{2n}\right) \le \frac{\text{area}(S_a^2)}{\text{area}(\text{cir}(S_a^2))} \le \frac{1}{3}\left(1 + \frac{1}{n}\right)\left(1 + \frac{1}{2n}\right) \tag{2.13}$$

for all $n \in \mathbf{Z}^+$.

The right side of (2.13) is greater than $\frac{1}{3}$ and the left side is less than $\frac{1}{3}$ for all $n \in \mathbf{Z}^+$, but by taking n large enough, both sides can be made as close to $\frac{1}{3}$ as we please. Hence we conclude that the ratio $\frac{\text{area}(S_a^2)}{\text{area}(\text{cir}(S_a^2))}$ is equal to $\frac{1}{3}$. Thus, we have proved the following theorem:

2.14 Theorem (Area Under a Parabola.) *Let a be a positive real number and let S_a^2 be the set of points (x, y) in $\mathbf{R}^2$ such that $0 \leq x \leq a$ and $0 \leq y \leq x^2$. Then*

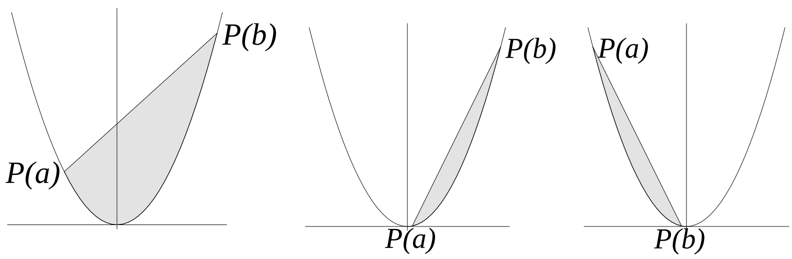
$$\frac{\text{area}(S_a^2)}{\text{area}(\text{box circumscribed about } S_a^2)} = \frac{1}{3},$$

i.e.

$$\text{area}(S_a^2) = \frac{1}{3}a^3.$$

Remark: The last paragraph of the proof of theorem 2.14 is a little bit vague. How large is "large enough" and what does "as close as we please" mean? Archimedes and Euclid would not have considered such an argument to be a proof. We will reconsider the end of this proof after we have developed the language to complete it more carefully. (Cf Example 6.54.)

The first person to calculate the area of a parabolic segment was Archimedes (287-212 B.C.). The parabolic segment considered by Archimedes corresponds to the set $S(a, b)$ bounded by the parabola $y = x^2$ and the line joining $P(a) = (a, a^2)$ to $P(b) = (b, b^2)$ where $(a < b)$.

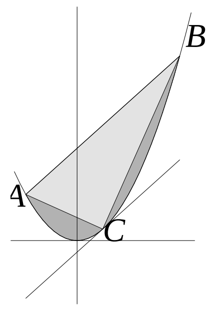


Parabolic Segments

2.15 Exercise. Show that the area of the parabolic segment $S(a,b)$ just described is $\frac{(b-a)^3}{6}$. Use theorem 2.14 and any results from Euclidean geometry that you need. You may assume that $0 < a \leq b$. The cases where $a < 0 < b$ and $a < b < 0$ are all handled by similar arguments.

The result of this exercise shows that the area of a parabolic segment depends only on its width. Thus the segment determined by the points $(-1,1)$ and $(1,1)$ has the same area as the segment determined by the points $(99,9801)$ and $(101,10201)$, even though the second segment is 400 times as tall as the first, and both segments have the same width. Does this seem reasonable?

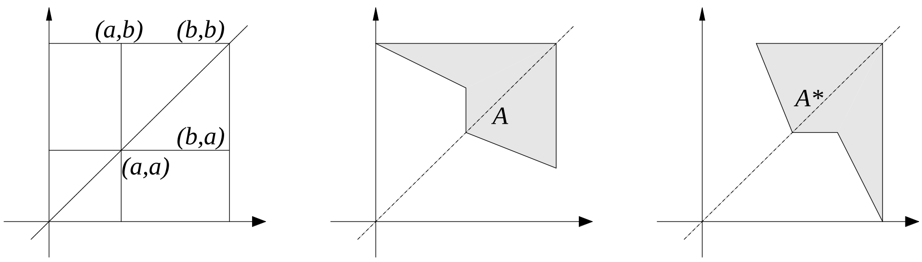
Remark: Archimedes stated his result about the area of a parabolic segment as follows. The area of the parabolic segment cut off by the line AB is four thirds of the area of the inscribed triangle ABC, where C is the point on the parabola at which the tangent line is parallel to AB. We cannot verify Archimedes formula at this time, because we do not know how to find the point C.



2.16 Exercise. Verify Archimedes' formula as stated in the above remark for the parabolic segment $S(-a,a)$. In this case you can use your intuition to find the tangent line.

The following definition is introduced as a hint for exercise 2.18

2.17 Definition (Reflection about the line $y = x$) If S is a subset of $\mathbf{R}^2$, then the *reflection of S about the line $y = x$* is defined to be the set of all points (x,y) such that $(y,x) \in S$.

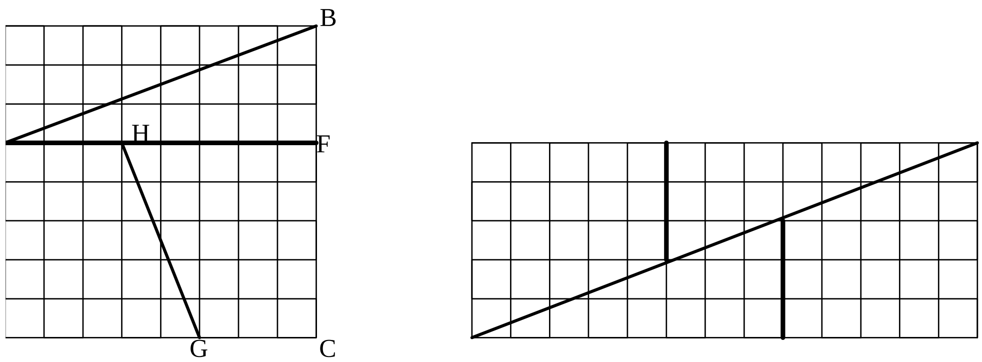


A^* is the reflection of A about the line $y = x$

If S^* denotes the reflection of S about the line $y = x$, then S and S^* have the same area.

2.18 Exercise. Let $a \in \mathbf{R}^+$ and let T_a be the set of all points (x, y) such that $0 \leq x \leq a$ and $0 \leq y \leq \sqrt{x}$. Sketch the set T_a and find its area.

2.19 Exercise. In the first figure below, the 8×8 square $ABCD$ has been divided into two 3×8 triangles and two trapezoids by means of the lines EF, EB and GH. In the second figure the four pieces have been rearranged to form an 5×13 rectangle. The square has area 64 , and the rectangle has area 65. Where did the extra unit of area come from? (This problem was taken from W. W. Rouse Ball's *Mathematical Recreations* [4, page 35]. Ball says that the earliest reference he could find for the problem is 1868.)



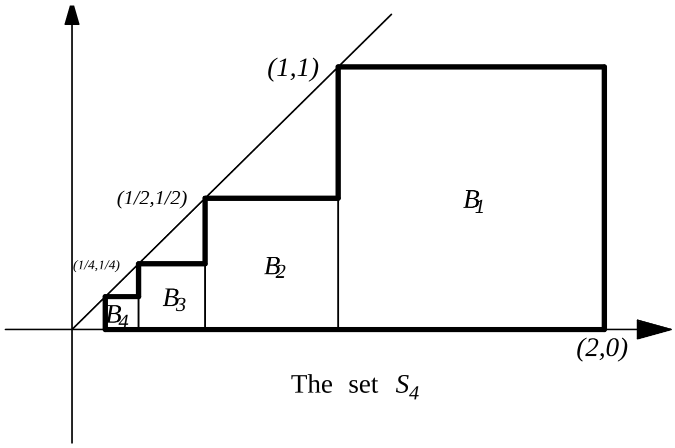
2.4 Finite Geometric Series

For each n in $\mathbf{Z}^+$ let B_n denote the box

$$B_n = B\left(\frac{1}{2^{n-1}}, \frac{2}{2^{n-1}} : 0, \frac{1}{2^{n-1}}\right),$$

and let

$$S_n = B_1 \cup B_2 \cup \cdots \cup B_n = \bigcup_{j=1}^{n} B_j.$$



The set S_4

I want to find the area of S_n. I have

$$\text{area}(B_n) = \left(\frac{2}{2^{n-1}} - \frac{1}{2^{n-1}}\right) \cdot \left(\frac{1}{2^{n-1}} - 0\right) = \frac{1}{2^{n-1}} \cdot \frac{1}{2^{n-1}} = \frac{1}{4^{n-1}}.$$

Since the boxes B_i intersect only along their boundaries, we have

$$\begin{aligned} \text{area}(S_n) &= \text{area}(B_1) + \text{area}(B_2) + \cdots + \text{area}(B_n) \\ &= 1 + \frac{1}{4} + \cdots + \frac{1}{4^{n-1}}. \end{aligned} \tag{2.20}$$

Thus

$$\begin{aligned} \text{area}(S_1) &= 1, \\ \text{area}(S_2) &= 1 + \frac{1}{4} = \frac{5}{4}, \\ \text{area}(S_3) &= \frac{5}{4} + \frac{1}{16} = \frac{20}{16} + \frac{1}{16} = \frac{21}{16} = \frac{21}{4^2}, \\ \text{area}(S_4) &= \frac{21}{16} + \frac{1}{64} = \frac{84}{64} + \frac{1}{64} = \frac{85}{64} = \frac{85}{4^3}. \end{aligned} \tag{2.21}$$

You probably do not see any pattern in the numerators of these fractions, but in fact area(S_n) is given by a simple formula, which we will now derive.

2.22 Theorem (Finite Geometric Series.) *Let r be a real number such that $r \neq 1$. Then for all $n \in \mathbf{Z}^+$*

$$1 + r + r^2 + \cdots + r^{n-1} = \frac{1 - r^n}{1 - r}. \tag{2.23}$$

Proof: Let

$$S = 1 + r + r^2 + \cdots + r^{n-1}.$$

Then

$$rS = r + r^2 + \cdots + r^{n-1} + r^n.$$

Subtract the second equation from the first to get

$$S(1 - r) = 1 - r^n,$$

and thus

$$S = \frac{1 - r^n}{1 - r}. \ \|\|[2]$$

Remark: Theorem 2.22 is very important, and you should remember it. Some people find it easier to remember the proof than to remember the formula. It would be good to remember both.

If we let $r = \frac{1}{4}$ in (2.23), then from equation (2.20) we obtain

$$\begin{aligned} \text{area}(S_n) &= 1 + \frac{1}{4} + \cdots + \frac{1}{4^{n-1}} \\ &= \frac{1 - \frac{1}{4^n}}{1 - \frac{1}{4}} = \frac{4}{3}\left(1 - \frac{1}{4^n}\right) \\ &= \frac{4^n - 1}{3 \cdot 4^{n-1}}. \end{aligned} \tag{2.24}$$

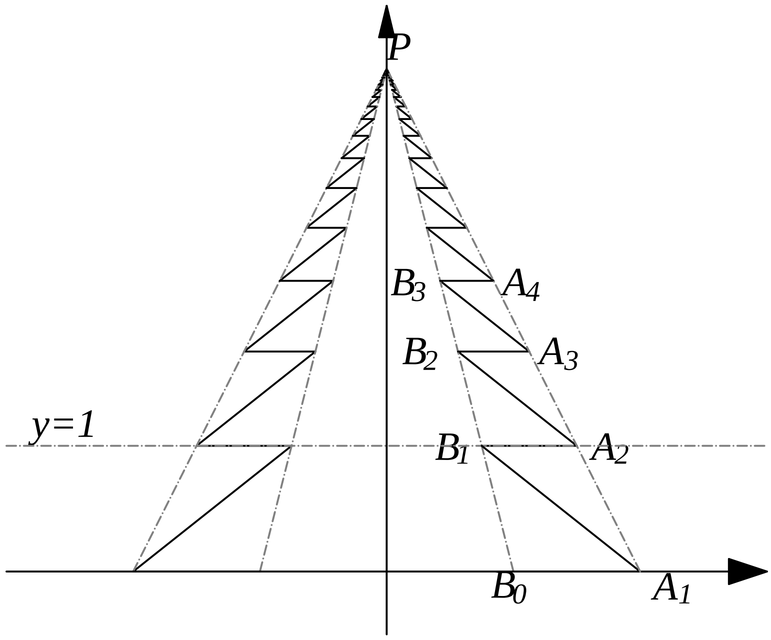
As a special case, we have

$$\text{area}(S_4) = \frac{4^4 - 1}{3 \cdot 4^3} = \frac{256 - 1}{3 \cdot 4^3} = \frac{255}{3 \cdot 4^3} = \frac{85}{4^3}$$

which agrees with equation (2.21).

[2] We use the symbol ||| to denote the end of a proof.

2.25 Entertainment (Pine Tree Problem.) Let T be the subset of $\mathbf{R}^2$ sketched below:



Here $P = (0,4)$, $B_0 = (1,0)$, $A_1 = (2,0)$, and B_1 is the point where the line B_0P intersects the line $y = 1$. All of the points A_j lie on the line PA_1, and all of the points B_j lie on the line PB_0. All of the segments A_iB_{i-1} are horizontal, and all segments A_jB_j are parallel to A_1B_1. Show that the area of T is $\frac{44}{7}$. You will probably need to use the formula for a geometric series.

2.26 Exercise.
(a) Find the number

$$1 + \frac{1}{7} + \frac{1}{7^2} + \frac{1}{7^3} + \cdots + \frac{1}{7^{100}}$$

accurate to 8 decimal places.
(b) Find the number

$$1 + \frac{1}{7} + \frac{1}{7^2} + \frac{1}{7^3} + \cdots + \frac{1}{7^{1000}}$$

accurate to 8 decimal places.

(You may use a calculator, but you can probably do this without using a calculator.)

2.27 Exercise. Let

$$\begin{aligned} a_1 &= .027 \\ a_2 &= .027027 \\ a_3 &= .027027027 \\ &\textit{etc.} \end{aligned}$$

Use the formula for a finite geometric series to get a simple formula for a_n. What rational number should the infinite decimal $.027027027\cdots$ represent? Note that

$$a_3 = .027(1.001001) = .027(1 + \frac{1}{1000} + \frac{1}{1000^2}).$$

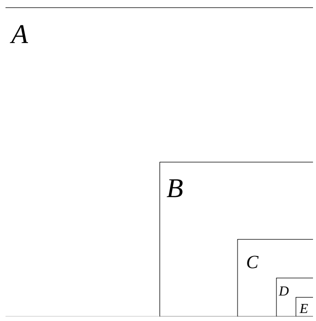
The Babylonians[45, page 77] knew that

$$1 + 2 + 2^2 + 2^3 + \cdots + 2^n = 2^n + (2^n - 1), \tag{2.28}$$

i.e. they knew the formula for a finite geometric series when $r = 2$.

Euclid knew a version of the formula for a finite geometric series in the case where r is a positive integer.

Archimedes knew the sum of the finite geometric series when $r = \frac{1}{4}$. The idea of Archimedes' proof is illustrated in the figure.



If the large square has side equal to 2, then

$$\begin{aligned} A & & &= A &= 3 \\ \tfrac{1}{4}A & & &= B & \\ (\tfrac{1}{4})^2 A &= \tfrac{1}{4}B & &= C & \\ (\tfrac{1}{4})^3 A &= \tfrac{1}{4}C & &= D. & \end{aligned}$$

Hence

$$\begin{aligned}(1+\tfrac{1}{4}+(\tfrac{1}{4})^2+(\tfrac{1}{4})^3)A &= (A+B+C+D) = 4-E \\ &= 4-(\tfrac{1}{8})^2 = 4-(\tfrac{1}{4})^3 = 4(1-(\tfrac{1}{4})^4).\end{aligned}$$

i.e.

$$(1+\tfrac{1}{4}+(\tfrac{1}{4})^2+(\tfrac{1}{4})^3)\cdot 3 = 4(1-(\tfrac{1}{4})^4).$$

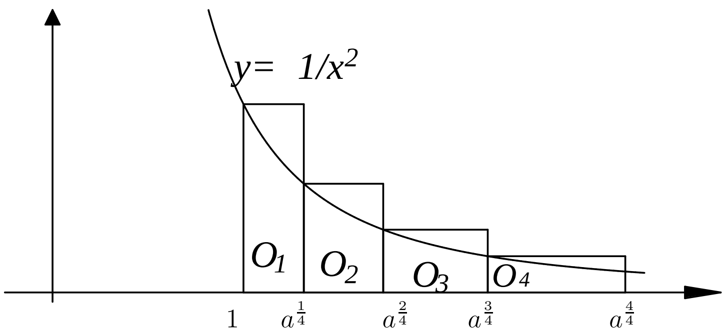
For the details of Archimedes' argument see [2, pages 249-250].

2.29 Exercise. Explain why formula (2.28) is a special case of the formula for a finite geometric series.

2.5 Area Under the Curve $y = \frac{1}{x^2}$

The following argument is due to Pierre de Fermat (1601-1665) [19, pages 219-222]. Later we will use Fermat's method to find the area under the curve $y = x^\alpha$ for all α in $\mathbf{R}\setminus\{-1\}$.

Let a be a real number with $a > 1$, and let S_a be the set of points (x, y) in $\mathbf{R}^2$ such that $1 \leq x \leq a$ and $0 \leq y \leq \dfrac{1}{x^2}$. I want to find the area of S_a.



Let n be a positive integer. Note that since $a > 1$, we have

$$1 < a^{\frac{1}{n}} < a^{\frac{2}{n}} < \cdots < a^{\frac{n}{n}} = a.$$

Let O_j be the box

$$O_j = B\left(a^{\frac{j-1}{n}}, a^{\frac{j}{n}} : 0, \frac{1}{\left(a^{\frac{j-1}{n}}\right)^2}\right).$$

Thus the upper left corner of O_j lies on the curve $y = \dfrac{1}{x^2}$.
To simplify the notation, I will write

$$b = a^{\frac{1}{n}}.$$

Then

$$O_j = B\left(b^{j-1}, b^j : 0, \frac{1}{b^{2(j-1)}}\right),$$

and

$$\text{area}(O_j) \;=\; \frac{b^j - b^{j-1}}{b^{2(j-1)}} = \frac{(b-1)b^{j-1}}{b^{2(j-1)}} = \frac{(b-1)}{b^{(j-1)}}.$$

Hence

$$\begin{aligned}
\text{area}\left(\bigcup_{j=1}^{n} O_j\right) &= \text{area}(O_1) + \text{area}(O_2) + \cdots + \text{area}(O_n) \\
&= (b-1) + \frac{(b-1)}{b} + \cdots + \frac{(b-1)}{b^{(n-1)}} \\
&= (b-1)\left(1 + \frac{1}{b} + \cdots + \frac{1}{b^{(n-1)}}\right).
\end{aligned}$$

Observe that we have here a finite geometric series, so

$$\text{area}\left(\bigcup_{j=1}^{n} O_j\right) = (b-1)\left(\frac{1-\frac{1}{b^n}}{1-\frac{1}{b}}\right) \tag{2.30}$$

$$= b\left(1-\frac{1}{b}\right)\left(\frac{1-\frac{1}{b^n}}{1-\frac{1}{b}}\right) = b\left(1-\frac{1}{b^n}\right). \tag{2.31}$$

Now

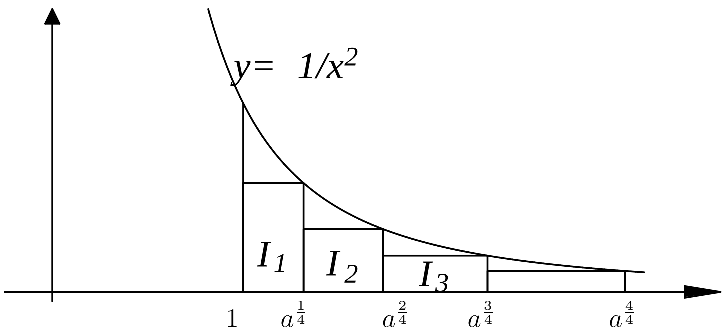
$$S_a \subset \bigcup_{j=1}^{n} O_j \tag{2.32}$$

so

$$\text{area}(S_a) \leq \text{area}(\bigcup_{j=1}^{n} O_j) = b\left(1 - \frac{1}{b^n}\right). \tag{2.33}$$

Let I_j be the box

$$I_j = B\left(a^{\frac{j-1}{n}}, a^{\frac{j}{n}} : 0, \frac{1}{a^{\frac{2j}{n}}}\right) = B\left(b^{j-1}, b^j : 0, \frac{1}{b^{2j}}\right)$$



so that the upper right corner of I_j lies on the curve $y = \frac{1}{x^2}$ and I_j lies underneath the curve $y = \frac{1}{x^2}$. Then

$$\begin{aligned} \text{area}(I_j) &= \left(\frac{b^j - b^{j-1}}{b^{2j}}\right) = \frac{(b-1)\,b^{j-1}}{b^{2j}} \\ &= \frac{(b-1)}{b^{(j+1)}} = \frac{(b-1)}{b^2 b^{j-1}} = \frac{\text{area}(O_j)}{b^2}. \end{aligned}$$

Hence,

$$\begin{aligned} \text{area}\left(\bigcup_{j=1}^{n} I_j\right) &= \text{area}(I_1) + \cdots + \text{area}(I_n) \\ &= \frac{\text{area}(O_1)}{b^2} + \cdots + \frac{\text{area}(O_n)}{b^2} = \frac{(\text{area}(O_1) + \cdots + \text{area}(O_n))}{b^2} \\ &= \frac{1}{b^2}\text{area}\left(\bigcup_{j=1}^{n} O_j\right) = \frac{1}{b^2} \cdot b(1 - \frac{1}{b^n}) = b^{-1}(1 - b^{-n}). \end{aligned}$$

Since

$$\bigcup_{j=1}^{n} I_j \subset S_a,$$

we have

$$\text{area}\left(\bigcup_{j=1}^{n} I_j\right) \le \text{area}(S_a);$$

i.e.,

$$b^{-1}(1-b^{-n}) \le \text{area}(S_a).$$

By combining this result with (2.33), we get

$$b^{-1}(1-b^{-n}) \le \text{area}(S_a) \le b(1-b^{-n}) \quad \text{for all} \quad n \in \mathbf{Z}^+.$$

Since $b = a^{\frac{1}{n}}$, we can rewrite this as

$$a^{-\frac{1}{n}}(1-a^{-1}) \le \text{area}(S_a) \le a^{\frac{1}{n}}(1-a^{-1}). \tag{2.34}$$

2.35 Exercise. What do you think the area of S_a should be? Explain your answer. If you have no ideas, take $a = 2$ in (2.34), take large values of n, and by using a calculator, estimate area(S_a) to three or four decimal places of accuracy.

2.36 Exercise. Let a be a real number with $0 < a < 1$, and let N be a positive integer. Then

$$a = a^{\frac{N}{N}} < a^{\frac{N-1}{N}} < \cdots < a^{\frac{2}{N}} < a^{\frac{1}{N}} < 1.$$

Let T_a be the set of points (x, y) such that $a \le x \le 1$ and $0 \le y \le \dfrac{1}{x^2}$. Draw a sketch of T_a, and show that

$$a^{\frac{1}{N}}(a^{-1}-1) \le \text{area}(T_a) \le a^{-\frac{1}{N}}(a^{-1}-1).$$

The calculation of area(T_a) is very similar to the calculation of area(S_a).

What do you think the area of T_a should be?

2.37 Exercise. Using the inequalities (2.6), and the results of Bernoulli's table on page 27, try to guess what the area of S_a^r is for an arbitrary positive integer r. Explain the basis for your guess. (The correct formula for area(S_a^r) for positive integers r was stated by Bonaventura Cavalieri in 1647[6, 122 ff]. Cavalieri also found a method for computing general positive integer power sums.)

2.6 *Area of a Snowflake.

In this section we will find the areas of two rather complicated sets, called the *inner snowflake* and the *outer snowflake*. To construct the inner snowflake, we first construct a family of polygons I_1, I_2, $I_3 \ldots$ as follows:

I_1 is an equilateral triangle.

I_2 is obtained from I_1 by adding an equilateral triangle to the middle third of each side of I_1, (see the figure on page 41).

I_3 is obtained from I_2 by adding an equilateral triangle to the middle third of each side of I_2, and in general

I_{n+1} is obtained from I_n by adding an equilateral triangle to the middle third of each side of I_n.

The inner snowflake is the set

$$K_I = \bigcup_{n=1}^{\infty} I_n,$$

i.e. a point is in the inner snowflake if and only if it lies in I_n for some positive integer n. Observe that the inner snowflake is not a polygon.

To construct the outer snowflake, we first construct a family of polygons O_1, O_2, $O_3 \ldots$ as follows:

O_1 is a regular hexagon.

O_2 is obtained from O_1 by removing an equilateral triangle from the middle third of each side of O_1, (see the figure on page 41).

O_3 is obtained from O_2 by removing an equilateral triangle from the middle third of each side of O_2, and in general

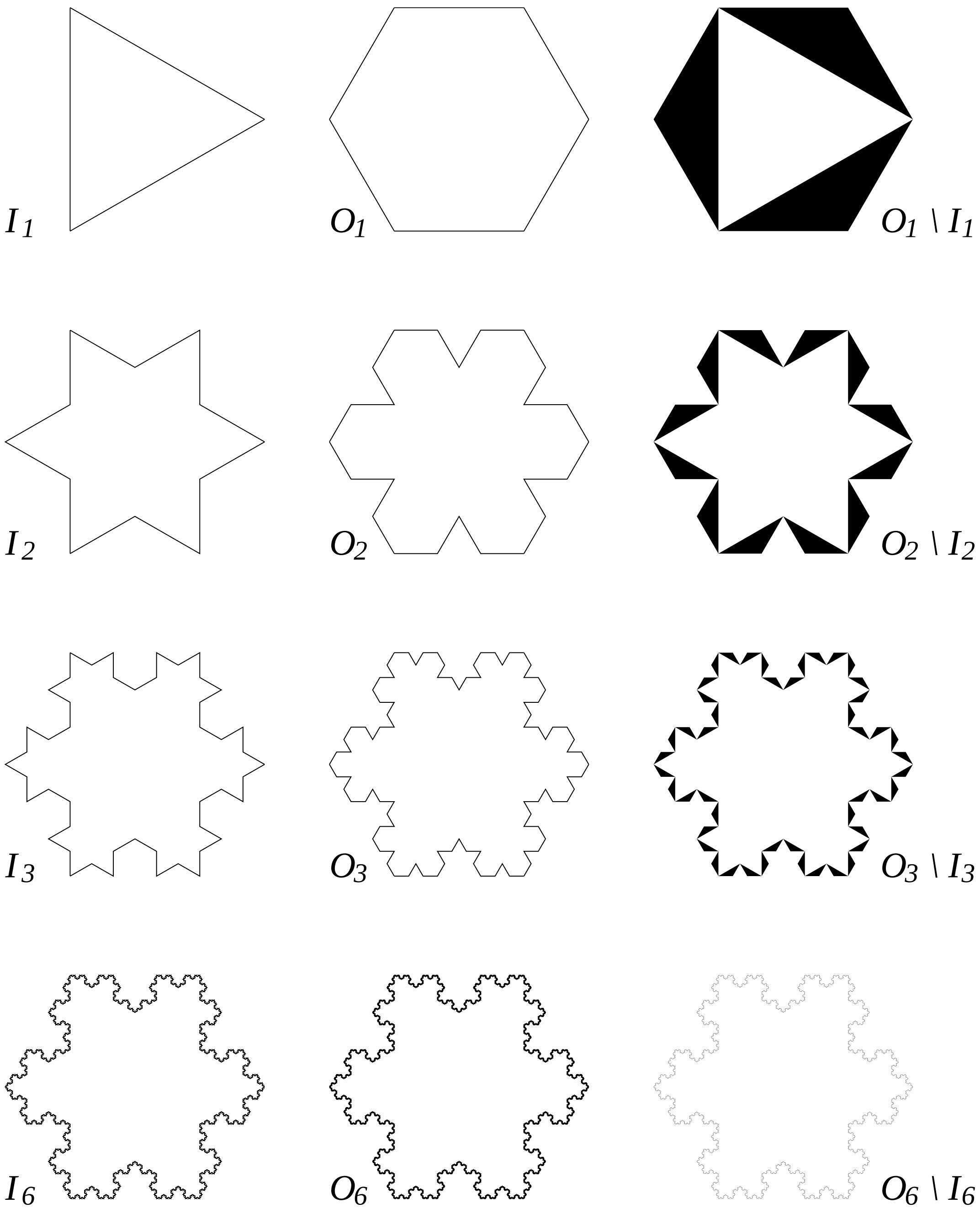
O_{n+1} is obtained from O_n by removing an equilateral triangle from the middle third of each side of O_n.

The outer snowflake is the set

$$K_O = \bigcap_{n=1}^{\infty} O_n,$$

i.e. a point is in the outer snowflake if and only if it lies in O_n for all positive integers n. Observe that the outer snowflake is not a polygon.

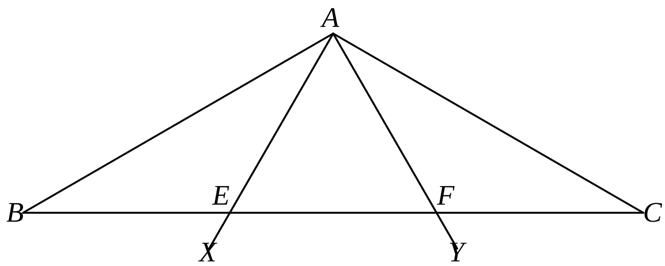
An isosceles 120° triangle is an isosceles triangle having a vertex angle of 120°. Since the sum of the angles of a triangle is two right angles, the base angles of such a triangle will be $\frac{1}{2}(180° - 120°) = 30°$.


I_1
O_1
$O_1 \setminus I_1$
I_2
O_2
$O_2 \setminus I_2$
I_3
O_3
$O_3 \setminus I_3$
I_6
O_6
$O_6 \setminus I_6$

Snowflakes

The following two technical lemmas[3] guarantee that in the process of building I_{n+1} from I_n we never reach a situation where two of the added triangles intersect each other, or where one of the added triangles intersects I_n, and in the process of building O_{n+1} from O_n we never reach a situation where two of the removed triangles intersect each other, or where one of the removed triangles fails to lie inside O_n.

2.38 Lemma. *Let $\triangle BAC$ be an isosceles 120° triangle with $\angle BAC = 120^\circ$. Let E, F be the points that trisect BC, as shown in the figure. Then $\triangle AEF$ is an equilateral triangle, and the two triangles $\triangle AEB$ and $\triangle AFC$ are congruent isosceles 120° triangles.*



Proof: Let $\triangle BAC$ be an isosceles triangle with $\angle BAC = 120^\circ$. Construct 30° angles BAX and CAY as shown in the figure, and let E and F denote the points where the lines AX and AY intersect BC. Then since the sum of the angles of a triangle is two right angles, we have

$$\angle AEB = 180^\circ - \angle ABE - \angle BAE = 180^\circ - 30^\circ - 30^\circ = 120^\circ.$$

Hence

$$\angle AEF = 180^\circ - \angle AEB = 180^\circ - 120^\circ = 60^\circ,$$

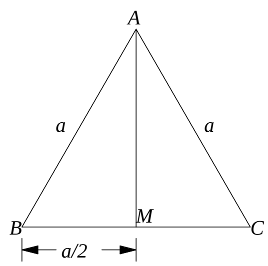
and similarly $\angle AFE = 60^\circ$. Thus $\triangle AEF$ is an isosceles triangle with two 60° angles, and thus $\triangle AEF$ is equilateral. Now $\angle BAE = 30^\circ$ by construction, and $\angle ABE = 30^\circ$ since $\angle ABE$ is a base angle of an isosceles 120° triangle. It follows that $\triangle BEA$ is isosceles and $BE = EA$. (If a triangle has two equal angles, then the sides opposite those angles are equal.) Thus, $BE = EA = EF$, and a similar argument shows that $CF = EF$. It follows that the points E and F

[3] A lemma is a theorem which is proved in order to help prove some other theorem.

trisect BC, and that $\triangle AEB$ is an isosceles 120° triangle. A similar argument shows that $\triangle AFC$ is an isosceles 120° triangle.

Now suppose we begin with the isosceles 120° triangle $\triangle BAC$ with angle $BAC = 120°$, and we let E, F be the points that trisect BC. Since A and E determine a unique line, it follows from the previous discussion that EA makes a 30° angle with BA and FA makes a 30° angle with AC, and that all the conclusions stated in the lemma are valid. |||

2.39 Lemma. *If T is an equilateral triangle with side of length a, then the altitude of T has length $\frac{a\sqrt{3}}{2}$, and the area of T is $\frac{\sqrt{3}}{4}a^2$. If R is an isosceles 120° triangle with two sides of length a, then the third side of R has length $a\sqrt{3}$.*



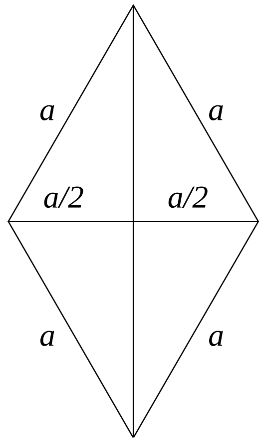
Proof: Let $T = \triangle ABC$ be an equilateral triangle with side of length a, and let M be the midpoint of BC. Then the altitude of T is AM, and by the Pythagorean theorem

$$AM = \sqrt{(AB)^2 - (BM)^2} = \sqrt{a^2 - \left(\frac{1}{2}a\right)^2} = \sqrt{\frac{3}{4}a^2} = \frac{\sqrt{3}}{2}a.$$

Hence

$$\text{area}(T) = \frac{1}{2}(\text{base})(\text{altitude}) = \frac{1}{2}a \cdot \frac{\sqrt{3}}{2}a = \frac{\sqrt{3}}{4}a^2.$$

An isosceles 120° triangle with two sides of length a can be constructed by taking halves of two equilateral triangles of side a, and joining them along their common side of length $\frac{a}{2}$, as indicated in the following figure.



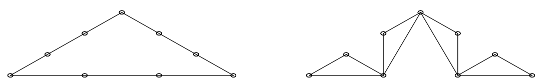
Hence the third side of an isosceles 120° triangle with two sides of length a is twice the altitude of an equilateral triangle of side a, i.e., is $2\left(\frac{\sqrt{3}}{2}a\right) = \sqrt{3}a$. ‖

We now construct two sequences of polygons. $I_1, I_2, I_3, \cdots$, and $O_1, O_2, O_3, \cdots$ such that

$$I_1 \subset I_2 \subset I_3 \subset \cdots \subset O_3 \subset O_2 \subset O_1.$$

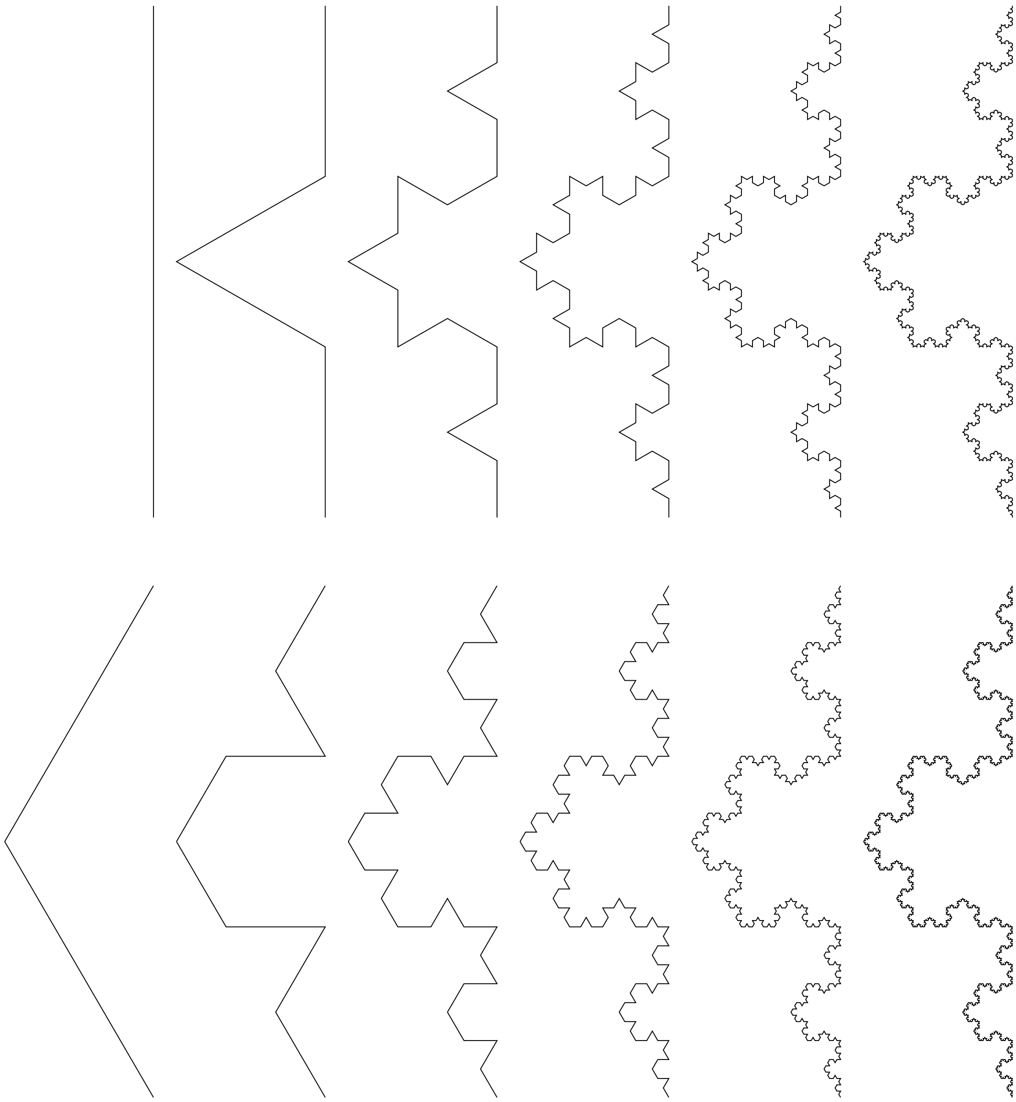
Let O_1 be a regular hexagon with side 1, and let I_1 be an equilateral triangle inscribed in O_1. Then $O_1 \setminus I_1$ consists of three isosceles 120° triangles with short side 1, and from lemma 2.39, it follows that the sides of I_1 have length $\sqrt{3}$. (See figure on page 41.)

Our general procedure for constructing polygons will be:



subtriangle of $O_n \setminus I_n$ $\rightarrow$ subtriangles of $O_{n+1} \setminus I_{n+1}$

O_{n+1} is constructed from O_n by removing an equilateral triangle from the middle third of each side of O_n, and I_{n+1} is constructed from I_n by adding an equilateral triangle to the middle third of each side of I_n. For each n, $O_n \setminus I_n$ will consist of a family of congruent isosceles 120° triangles and $O_{n+1} \setminus I_{n+1}$ is obtained from $O_n \setminus I_n$ by removing an equilateral triangle from the middle third of each side of each isosceles 120° triangle. Pictures of I_n, O_n, and $O_n \setminus I_n$ are given on page 41. Details of the pictures are shown on page 45.



Details of snowflakes

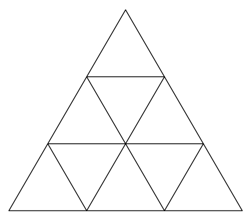
Lemma 2.38 guarantees that this process always leads from a set of isosceles 120° triangles to a new set of isosceles 120° triangles. Note that every vertex of O_n is a vertex of O_{n+1} and of I_{n+1}, and every vertex of I_n is a vertex of O_n and of I_{n+1}.
Let

$$\begin{aligned}
s_n &= \text{length of a side of } I_n.\\
t_n &= \text{area of equilateral triangle with side } s_n.\\
m_n &= \text{number of sides of } I_n.\\
a_n &= \text{area of } I_n.\\
S_n &= \text{length of a side of } O_n.\\
T_n &= \text{area of equilateral triangle with side } S_n.\\
M_n &= \text{number of sides of } O_n.\\
A_n &= \text{area of } O_n.
\end{aligned}$$

Then

$$\begin{array}{ll}
s_{n+1} = \frac{1}{3}s_n, & S_{n+1} = \frac{1}{3}S_n,\\
m_{n+1} = 4m_n, & M_{n+1} = 4M_n,\\
a_{n+1} = a_n + m_n t_{n+1} & A_{n+1} = A_n - M_n T_{n+1}.
\end{array}$$

Since an equilateral triangle with side s can be decomposed into nine equilateral triangles of side $\frac{s}{3}$ (see the figure),



we have

$$t_{n+1} = \frac{t_n}{9} \text{ and } T_{n+1} = \frac{T_n}{9}.$$

Also

$$a_1 = \text{area}(I_1) = t_1,$$

and since O_1 can be written as a union of six equilateral triangles,

$$A_1 = 6T_1.$$

The following table summarizes the values of s_n, m_n, t_n, S_n, M_n and T_n:

n	m_n	t_n	$m_{n-1}t_n$	M_n	T_n	$M_{n-1}T_n$
1	3	a_1		6	$\frac{A_1}{6}$	
2	$3 \cdot 4$	$\frac{a_1}{9}$	$\frac{3}{9}a_1$	$6 \cdot 4$	$\frac{1}{9}\frac{A_1}{6}$	$\frac{A_1}{9}$
3	$3 \cdot 4^2$	$\frac{a_1}{9^2}$	$\frac{3}{9} \cdot \frac{4}{9}a_1$	$6 \cdot 4^2$	$\frac{1}{9^2}\frac{A_1}{6}$	$\frac{4}{9}\frac{A_1}{9}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$3 \cdot 4^{n-1}$	$\frac{a_1}{9^{n-1}}$	$\frac{3}{9}\left(\frac{4}{9}\right)^{n-2} a_1$	$6 \cdot 4^{n-1}$	$\frac{1}{9^{n-1}}\frac{A_1}{6}$	$\left(\frac{4}{9}\right)^{n-2}\frac{A_1}{9}$

Now

$$\begin{aligned}
A_2 &= A_1 - M_1T_2 = A_1 - \frac{A_1}{9}, \\
A_3 &= A_2 - M_2T_3 = A_1 - \frac{A_1}{9} - \left(\frac{4}{9}\right)\frac{A_1}{9}, \\
&\vdots \\
A_{n+1} &= A_n - M_nT_{n+1} \\
&= A_1 - \frac{A_1}{9} - \left(\frac{4}{9}\right)\frac{A_1}{9} - \left(\frac{4}{9}\right)^2\frac{A_1}{9} - \cdots - \left(\frac{4}{9}\right)^{n-1}\frac{A_1}{9} \\
&= A_1 - \frac{A_1}{9}\left(1 + \frac{4}{9} + \left(\frac{4}{9}\right)^2 + \cdots + \left(\frac{4}{9}\right)^{n-1}\right). \qquad (2.40)
\end{aligned}$$

Also,

$$\begin{aligned}
a_2 &= a_1 + m_1t_2 = a_1 + \frac{3}{9}a_1, \\
a_3 &= a_2 + m_2t_3 = a_1 + \frac{3}{9}a_1 + \frac{3}{9}\left(\frac{4}{9}\right)a_1, \\
&\vdots \\
a_{n+1} &= a_n + m_nt_{n+1} \\
&= a_1 + \frac{3}{9}a_1 + \frac{3}{9}\left(\frac{4}{9}\right)a_1 + \frac{3}{9}\left(\frac{4}{9}\right)^2 a_1 + \cdots + \frac{3}{9}\left(\frac{4}{9}\right)^{n-1} a_1 \\
&= a_1 + \frac{a_1}{3}\left(1 + \frac{4}{9} + \left(\frac{4}{9}\right)^2 + \cdots + \left(\frac{4}{9}\right)^{n-1}\right). \qquad (2.41)
\end{aligned}$$

By the formula for a finite geometric series we have

$$1+\frac{4}{9}+(\frac{4}{9})+\cdots+(\frac{4}{9})^{n-1}=\frac{1-(\frac{4}{9})^n}{1-\frac{4}{9}}=\frac{9}{5}\left[1-(\frac{4}{9})^n\right].$$

By using this result in equations (2.40) and (2.41) we obtain

$$\begin{aligned}\text{area}(O_{n+1}) &= A_{n+1}=A_1-\frac{A_1}{5}\left[1-\left(\frac{4}{9}\right)^n\right]\\ &= \frac{4}{5}A_1+\frac{A_1}{5}\left(\frac{4}{9}\right)^n, \end{aligned} \tag{2.42}$$

and

$$\begin{aligned}\text{area}(I_{n+1}) &= a_{n+1}=a_1+\frac{a_1}{3}\cdot\frac{9}{5}\left[1-\left(\frac{4}{9}\right)^n\right]\\ &= \frac{8}{5}a_1-\frac{3a_1}{5}\left(\frac{4}{9}\right)^n. \end{aligned}$$

Now you can show that $a_1=\frac{A_1}{2}$, so the last equation may be written as

$$\text{area}(I_{n+1})=\frac{4}{5}A_1-\frac{3a_1}{5}\left(\frac{4}{9}\right)^n. \tag{2.43}$$

2.44 Exercise. Show that $a_1=\frac{A_1}{2}$, i.e. show that $\text{area}(I_1)=\frac{1}{2}\text{area}(O_1)$.

2.45 Definition (Snowflakes.) Let $K_I=\bigcup_{n=1}^{\infty} I_n$ and $K_O=\bigcap_{n=1}^{\infty} O_n$. Here the *infinite union* $\bigcup_{n=1}^{\infty} I_n$ means the set of all points x such that $x\in I_n$ for some n in $\mathbf{Z}^+$, and the *infinite intersection* $\bigcap_{n=1}^{\infty} O_n$ means the set of points x that are in all of the sets O_n where $n\in\mathbf{Z}^+$. I will call the sets K_I and K_O the *inner snowflake* and the *outer snowflake*, respectively.

For all k in $\mathbf{Z}^+$, we have

$$I_k\subset\bigcup_{n=1}^{\infty} I_n=K_I\subset K_O=\bigcap_{n=1}^{\infty} O_n\subset O_k,$$

so

$$\text{area}(I_k) \leq \text{area}(K_I) \leq \text{area}(K_O) \leq \text{area}(O_k).$$

Since $\left(\frac{4}{9}\right)^n$ can be made very small by taking n large (see theorem 6.71), we conclude from equations 2.43 and 2.42 that

$$\text{area}(K_I) = \text{area}(K_O) = \frac{4}{5}A_1 = \frac{4}{5}\text{area}(O_1).$$

We will call O_1 the *circumscribed hexagon* for K_I and for K_O. We have proved the following theorem:

2.46 Theorem. *The area of the inner snowflake and the outer snowflake are both $\frac{4}{5}$ of the area of the circumscribed hexagon.*

Note that both snowflakes touch the boundary of the circumscribed hexagon in infinitely many points.

It is natural to ask whether the sets K_O and K_I are the same.

2.47 Entertainment (Snowflake Problem.) Show that the inner snowflake is not equal to the outer snowflake. In fact, there are points in the boundary of the circumscribed hexagon that are in the outer snowflake but not in the inner snowflake.

The snowflakes were discovered by Helge von Koch(1870–1924), who published his results in 1906 [31]. Actually Koch was not interested in the snowflakes as two-dimensional objects, but as one-dimensional curves. He considered only part of the boundary of the regions we have described. He showed that the boundary of K_O and K_I is a curve that does not have a tangent at any point. You should think about the question: "In what sense is the boundary of K_O a curve?" In order to answer this question you would need to answer the questions "what is a curve?" and "what is the boundary of a set in $\mathbf{R}^2$?" We will not consider these questions in this course, but you might want to think about them.

I will leave the problem of calculating the perimeter of a snowflake as an exercise. It is considerably easier than finding the area.

2.48 Exercise. Let I_n and O_n be the polygons described in section 2.6, which are contained inside and outside of the snowflakes K_I and K_O.

a) Calculate the length of the perimeter of I_n.

b) Calculate the length of the perimeter of O_n.

What do you think the perimeter of K_O should be? (Since it isn't really clear what we mean by "the perimeter of K_O," this question doesn't really have a "correct" answer – but you should come up with some answer.)

Chapter 3

Propositions and Functions

In this chapter we will introduce some general mathematical ideas and notation that will be useful in the following chapters.

3.1 Propositions

3.1 Definition (Proposition.) A *proposition* is a statement that is either true or false. I will sometimes write a proposition inside of quotes (“ ”), when I want to emphasize where the proposition begins and ends.

3.2 Examples.
If $P_1 =$ “$1 + 1 = 2$”, then P_1 is a true proposition.
If $P_2 =$ “$1 + 1 = 3$”, then P_2 is a false proposition.
If $P_3 =$ “2 is an even number”, then P_3 is a true proposition.
If $P_4 =$ “7 is a lucky number”, then I will not consider P_4 to be a proposition (unless *lucky number* has been defined.)

3.3 Definition (And, or, not.) Suppose that P and Q are propositions. Then we can form new propositions denoted by “P and Q”, “P or Q”, and “not P”.
“P and Q” is true if and only if both of P, Q are true.
“P or Q” is true if and only if at least one of P, Q is true.
“not P” is true if and only if P is false.
Observe that in mathematics, “or” is always assumed to be inclusive or: If “P” and “Q” are both true, then “P or Q” is true.

3.4 Examples.
"$1 + 1 = 2$ and $1 + 1 = 3$" is false.
"$1 + 1 = 2$ or $1 + 1 = 3$" is true.
"$1 + 1 = 2$ or $2 + 2 = 4$" is true.
"not(not P)" is true if and only if P is true.

For each element x of $\mathbf{Q}$ let $R(x)$ be the proposition "$x^2 + 5x + 6 = 0$". Thus $R(-3) =$ "$(-3)^2 + 5 \cdot (-3) + 6 = 0$", so $R(-3)$ is true, while $R(0) =$ "$0^2 + 5 \cdot 0 + 6 = 0$", so $R(0)$ is false. Here I consider R to be a rule which assigns to each element x of $\mathbf{Q}$ a proposition $R(x)$.

3.5 Definition (Proposition form.) Let S be a set. A rule P that assigns to each element x of S a unique proposition $P(x)$ is called a *proposition form over* S.

Thus the rule R defined in the previous paragraph is a proposition form over $\mathbf{Q}$. Note that a proposition form is neither true nor false, i.e. a proposition form is not a proposition.

3.6 Definition ($\Longleftrightarrow$, Equivalent propositions.) Let P, Q be two propositions. We say that "P is *equivalent to* Q" if either (P, Q are both true) or (P, Q are both false). Thus every proposition is equivalent either to "$1+1 = 2$" or to "$1+1 = 3$." We write "$P \Longleftrightarrow Q$" as an abbreviation for "P is equivalent to Q." If P, Q are propositions, then "$P \Longleftrightarrow Q$" is a proposition, and
"$P \Longleftrightarrow Q$" is true if and only if ((P, Q are both true) or (P, Q are both false)).

Ordinarily one would not make a statement like
"$(1 + 1 = 2) \Longleftrightarrow$ (4421 is a prime number)"
even though this is a true proposition. One writes "$P \Longleftrightarrow Q$" in an argument, only when the person reading the argument can be expected to see the equivalence of the two statements P and Q.

If P, Q, R and S are propositions,then

$$P \Longleftrightarrow Q \Longleftrightarrow R \Longleftrightarrow S \tag{3.7}$$

is an abbreviation for

$$((P \Longleftrightarrow Q) \text{ and } (Q \Longleftrightarrow R)) \text{ and } (R \Longleftrightarrow S).$$

Thus if we know that (3.7) is true, then we can conclude that $P \Longleftrightarrow S$ is true. The statement "$P \Longleftrightarrow Q$" is sometimes read as " P if and only if Q".

3.8 Example. Find all real numbers x such that

$$x^2 - 5x + 6 = 0. \tag{3.9}$$

Let x be an arbitrary real number. Then

$$\begin{aligned} x^2 - 5x + 6 = 0 &\iff (x-2)(x-3) = 0 \\ &\iff ((x-2) = 0) \text{ or } ((x-3) = 0) \\ &\iff (x = 2) \text{ or } (x = 3). \end{aligned}$$

Thus the set of all numbers that satisfy equation (3.9) is $\{2,3\}$. ‖

3.10 Definition ($\Longrightarrow$, Implication.) If P and Q are propositions then we say "P implies Q" and write "$P \Longrightarrow Q$", if the truth of Q follows from the truth of P. We make the convention that if P is false then $(P \Longrightarrow Q)$ is true for all propositions Q, and in fact that

$$(P \Longrightarrow Q) \text{ is true } \textit{unless } (P \text{ is true and } Q \text{ is false}). \tag{3.11}$$

Hence for all propositions P and Q

$$(P \Longrightarrow Q) \iff (Q \text{ or } \operatorname{not}(P)). \tag{3.12}$$

3.13 Example. For every element x in $\mathbf{Q}$

$$x = 2 \Longrightarrow x^2 = 4. \tag{3.14}$$

In particular, the following statements are all true.

$$2 = 2 \Longrightarrow 2^2 = 4. \tag{3.15}$$

$$-2 = 2 \Longrightarrow (-2)^2 = 4. \tag{3.16}$$

$$3 = 2 \Longrightarrow 3^2 = 4. \tag{3.17}$$

In proposition 3.16, P is false, Q is true, and $P \Longrightarrow Q$ is true.
In proposition 3.17, P is false, Q is false, and $P \Longrightarrow Q$ is true.

The usual way to prove $P \Longrightarrow Q$ is to assume that P is true and show that then Q must be true. This is sufficient by our convention in (3.11).

If P and Q are propositions, then "$P \Longrightarrow Q$" is also a proposition, and

$$(P \Longleftrightarrow Q) \text{ is equivalent to } (P \Longrightarrow Q \text{ and } Q \Longrightarrow P) \tag{3.18}$$

(the right side of (3.18) is true if and only if P, Q are both true or both false.) An alternate way of writing "$P \Longrightarrow Q$" is "if P then Q".

We will not make much use of the idea of two propositions being equal. Roughly, two propositions are equal if and only if they are word for word the same. Thus "$1 + 1 = 2$" and "$2 = 1 + 1$" are not equal propositions, although they are equivalent. The only time I will use an "=" sign between propositions is in definitions. For example, I might define a proposition form P over $\mathbf{N}$ by saying

for all $n \in \mathbf{N}$, $P(n) =$"$n + 1 = 2$",

or

for all $n \in \mathbf{N}$, $P(n) = [n + 1 = 2]$.

The definition we have given for "implies" is a matter of convention, and there is a school of contemporary mathematicians (called constructivists) who define $P \Longrightarrow Q$ to be true only if a "constructive" argument can be given that the truth of Q follows from the truth of P. For the constructivists, some of the propositions of the sort we use are neither true nor false, and some of the theorems we prove are not provable (or disprovable). A very readable description of the constructivist point of view can be found in the article *Schizophrenia in Contemporary Mathematics*[10, pages 1–10].

3.19 Exercise.

a) Give examples of propositions P, Q such that "$P \Longrightarrow Q$" and "$Q \Longrightarrow P$" are both true, or else explain why no such examples exist.

b) Give examples of propositions R, S such that "$R \Longrightarrow S$" and "$S \Longrightarrow R$" are both false, or explain why no such examples exist.

c) Give examples of propositions T, V such that "$T \Longrightarrow V$" is true but "$V \Longrightarrow T$" is false, or explain why no such examples exist.

3.20 Exercise. Let P, Q be two propositions. Show that the propositions "$P \Longrightarrow Q$" and " not$Q \Longrightarrow$ not$P$" are equivalent. ("not$Q \Longrightarrow$ not$P$" is called the *contrapositive* of the statement "$P \Longrightarrow Q$".)

3.21 Exercise. Which of the proposition forms below are true for all real numbers x? If a proposition form is not true for all real numbers x, give a number for which it is false.

a) $x = 1 \Longrightarrow x^2 = 1$.

b) $x^2 = 1 \implies x = 1$.

c) $x < \frac{1}{2} \implies 2x < 1$.

d) $2 < \frac{1}{x} \iff 2x < 1$. (Here assume $x \neq 0$.)

e) $x < 1 \implies x + 1 < 3$.

f) $x < 1 \iff x + 1 < 3$.

g) $x \leq 1 \implies x < 1$.

h) $x < 1 \implies x \leq 1$.

3.22 Exercise. Both of the arguments A and B given below are faulty, although one of them leads to a correct conclusion. Criticize both arguments, and correct one of them.

Problem: Let S be the set of all real numbers x such that $x \neq -2$. Describe the set of all elements $x \in S$ such that

$$\frac{12}{x+2} < 4. \tag{3.23}$$

Note that if $x \in S$ then $\frac{12}{x+2}$ is defined.

ARGUMENT A: Let x be an arbitrary element of S. Then

$$\begin{aligned} \frac{12}{x+2} < 4 &\iff 12 < 4x + 8 \\ &\iff 0 < 4x - 4 \\ &\iff 0 < 4(x-1) \\ &\iff 0 < x - 1 \\ &\iff 1 < x. \end{aligned}$$

Hence the set of all real numbers that satisfy inequality (3.23) is the set of all real numbers x such that $1 < x$. ‖

ARGUMENT B: Let x be an arbitrary element of S. Then

$$\begin{aligned} \frac{12}{x+2} < 4 &\implies 0 < 4 - \frac{12}{x+2} \\ &\implies 0 < \frac{4x + 8 - 12}{x+2} \end{aligned}$$

$$\begin{aligned}
&\Longrightarrow \quad 0 < \frac{4x-4}{x+2} \\
&\Longrightarrow \quad 0 < \frac{4(x-1)}{x+2} \\
&\Longrightarrow \quad 0 < \frac{x-1}{x+2}.
\end{aligned}$$

Now

$$\begin{aligned}
0 < \frac{x-1}{x+2} \quad &\Longleftrightarrow \quad (0 < x-1 \text{ and } \ 0 < x+2) \text{ or } \ (0 > x-1 \text{ and } \ 0 > x+2) \\
&\Longleftrightarrow \quad (1 < x \text{ and } \ -2 < x) \text{ or } \ (1 > x \text{ and } \ -2 > x) \\
&\Longleftrightarrow \quad 1 < x \text{ or } \ -2 > x.
\end{aligned}$$

Hence the set of all real numbers that satisfy inequality (3.23) is the set of all $x \in \mathbf{R}$ such that either $x < -2$ or $x > 1$. |||

3.2 Sets Defined by Propositions

The most common way of describing sets is by means of proposition forms.

3.24 Notation ($\{x : P(x)\}$) Let P be a proposition form over a set S, and let T be a subset of S. Then

$$\{x : x \in T \text{ and } \ P(x)\} \tag{3.25}$$

is defined to be the set of all elements x in T such that $P(x)$ is true. The set described in (3.25) is also written

$$\{x \in T : P(x)\}.$$

In cases where the meaning of "T" is clear from the context, we may abbreviate (3.25) by

$$\{x : P(x)\}.$$

3.26 Examples.

$$\{x \in \mathbf{Z} \ : \ \text{for some } y \in \mathbf{Z} \ (x = 2y)\}$$

is the set of all even integers, and

$$\mathbf{Z}^+ \ = \{x : x \in \mathbf{Z} \ \text{ and } \ x > 0\}.$$

If A and B are sets, then

$$A \cup B = \{x : x \in A \text{ or } x \in B\}, \tag{3.27}$$
$$A \cap B = \{x : x \in A \text{ and } x \in B\}, \tag{3.28}$$
$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}. \tag{3.29}$$

We will use the following notation throughout these notes.

3.30 Notation ($\mathbf{Z}_{\geq n}$, $\mathbf{R}_{\geq a}$) If n is an integer we define

$$\mathbf{Z}_{\geq n} = \{k \in \mathbf{Z} : k \geq n\}.$$

Thus

$$\mathbf{Z}_{\geq 1} = \mathbf{Z}^{+} \text{ and } \mathbf{Z}_{\geq 0} = \text{the set of non-negative integers} = \mathbf{N}.$$

Similarly, if a is a real number, we define

$$\mathbf{R}_{\geq a} = \{x \in \mathbf{R} : x \geq a\}.$$

3.31 Definition (Ordered pair.) If a, b are objects, then the *ordered pair* (a, b) is a new object obtained by combining a and b. Two ordered pairs (a, b) and (c, d) are equal if and only if $a = c$ and $b = d$. Similarly we may consider *ordered triples*. Two ordered triples (a, b, x) and (c, d, y) are equal if and only if $a = c$ and $b = d$ and $x = y$. We use the same notation (a, b) to represent an open interval in $\mathbf{R}$ and an ordered pair in $\mathbf{R}^2$. The context should always make it clear which meaning is intended.

3.32 Definition (Cartesian product) If A, B are sets then the *Cartesian product of A and B* is defined to be the set of all ordered pairs (x, y) such that $x \in A$ and $y \in B$:

$$A \times B = \{(x, y) : x \in A \text{ and } y \in B\} \tag{3.33}$$

3.34 Examples. Let a, b, c, d be real numbers with $a \leq b$ and $c \leq d$. Then

$$[a, b] \times [c, d] = B(a, b : c, d)$$

and

$$[c, d] \times [a, b] = B(c, d : a, b).$$

Thus in general $A \times B \neq B \times A$.

The set $A \times A$ is denoted by A^2. You are familiar with one Cartesian product. The euclidean plane $\mathbf{R}^2$ is the Cartesian product of $\mathbf{R}$ with itself.

3.35 Exercise. Let $S = B(-2,2:-2,2)$ and let
$R_1 = \{(x,y) \in S : xy \leq 0\}$
$R_2 = \{(x,y) \in S : x^2 - 1 \leq 0\}$
$R_3 = \{(x,y) \in S : y^2 - 1 \leq 0\}$
$R_4 = \{(x,y) \in S : xy(x^2-1)(y^2-1) \leq 0\}$
Sketch the sets S, R_1, R_2, R_3, R_4. For R_4 you should include an explanation of how you arrived at your answer. For the other sets no explanation is required.

3.36 Exercise. Do there exist sets A, B such that $A \times B$ has exactly five elements?

3.3 Functions

3.37 Definition (Function.) Let A, B be sets. A *function with domain A and codomain B* is an ordered triple (A, B, f), where f is a rule which assigns to each element of A a unique element of B. The element of B which f assigns to an element x of A is denoted by $f(x)$. We call $f(x)$ the *f-image of x* or the *image of x under f*. The notation $f : A \longrightarrow B$ is an abbreviation for "f is a function with domain A and codomain B". We read "$f : A \longrightarrow B$" as "f is a function from A to B."

3.38 Examples. Let $f : \mathbf{Z} \longrightarrow \mathbf{N}$ be defined by the rule

$$f(n) = n^2 \text{ for all } n \in \mathbf{Z}.$$

Then $f(2) = 4, f(-2) = 4,$ and $f(1/2)$ is not defined, because $1/2 \notin \mathbf{Z}$.

Let $g : \mathbf{N} \longrightarrow \mathbf{N}$ be defined by the rule: for all $n \in \mathbf{N}$

$$g(n) = \text{the last digit in the decimal expansion for } n.$$

Thus $g(21) = 1$, $g(0) = 0$, $g(1984) = 4$, $g(666) = 6$.

3.39 Definition (Maximum and minimum functions.) We define functions max and min from $\mathbf{R}^2$ to $\mathbf{R}$ by the rule

$$\max(x,y) = \begin{cases} x & \text{if } x \geq y \\ y & \text{otherwise.} \end{cases} \tag{3.40}$$

$$\min(x,y) = \begin{cases} y & \text{if } x \geq y \\ x & \text{otherwise.} \end{cases} \tag{3.41}$$

Thus we have

$$\min(x,y) \leq x \leq \max(x,y)$$

and

$$\min(x,y) \leq y \leq \max(x,y)$$

for all $(x,y) \in \mathbf{R}^2$. Also

$$\max(2,7) = 7 \text{ and } \min(-2,-7) = -7.$$

3.42 Definition (Absolute value function.) Let $A\colon \mathbf{R} \to \mathbf{R}$ be defined by the rule

$$A(x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x < 0. \end{cases}$$

We call A the *absolute value function* and we usually designate $A(x)$ by $|x|$.

3.43 Definition (Sequence) Let S be a set. A *sequence* in S is a function $f\colon \mathbf{Z}^+ \to S$. I will refer to a sequence in $\mathbf{R}$ as a *real sequence.*

The sequence f is sometimes denoted by $\{f(n)\}$. Thus $\left\{\frac{1}{n^2+1}\right\}$ is the sequence $f\colon \mathbf{Z}^+ \to \mathbf{R}$ such that $f(n) = \frac{1}{n^2+1}$ for all $n \in \mathbf{Z}^+$. Sometimes the sequence f is denoted by

$$\{f(1), f(2), f(3), \cdots\}, \tag{3.44}$$

for example $\{1, \frac{1}{2}, \frac{1}{3}, \cdots\}$ is the same as $\{\frac{1}{n}\}$. The notation in formula (3.44) is always ambiguous. I will use it for sequences like

$$\{1, 1, -1, -1, 1, 1, -1, -1, 1, 1 \cdots\}$$

in which it is somewhat complicated to give an analytic description for $f(n)$.

If f is a sequence, and $n \in \mathbf{Z}^+$, then we often denote $f(n)$ by f_n.

3.45 Examples. Let P denote the set of all polygons in the plane. For each number a in $\mathbf{R}^+$ let

$$S_a^2 = \{(x,y) \in \mathbf{R}^2 : 0 \leq x \leq a \text{ and } 0 \leq y \leq x^2\}.$$

For each $n \in \mathbf{Z}^+$ let

$$Q_n = \bigcup_{i=1}^{n} I_i$$

and

$$R_n = \bigcup_{i=1}^{n} O_i$$

denote the polygons inscribed in S_a^2 and containing S_a^2 described on page 20. Then

$\{Q_n\}$ and $\{R_n\}$ are sequences in P.
$\{\text{area}(Q_n)\} = \{\frac{a^3}{3}(1-\frac{1}{n})(1-\frac{1}{2n})\}$ is a real sequence. (Cf. (2.3) and (2.12).)
$\{[\text{area}(Q_n), \text{area}(R_n)]\}$ is a sequence of intervals.

3.46 Definition (Equality for functions.) Let (A, B, f) and (C, D, g) be two functions. Then, since a function is an ordered triple, we have

$$(A, B, f) = (C, D, g) \text{ if and only if } A = C \text{ and } B = D, \text{ and } f = g.$$

The rules f and g are equal if and only if $f(a) = g(a)$ for all $a \in A$. If $f : A \longrightarrow B$ and $g : C \longrightarrow D$ then it is customary to write $f = g$ to mean $(A, B, f) = (C, D, g)$. This is an abuse of notation, but it is a standard practice.

3.47 Examples. If $f : \mathbf{Z} \longrightarrow \mathbf{Z}$ is defined by the rule

$$f(x) = x^2 \text{ for all } x \text{ in } \mathbf{Z}$$

and $g : \mathbf{Z} \longrightarrow \mathbf{N}$ is defined by the rule

$$g(x) = x^2 \text{ for all } x \text{ in } \mathbf{Z}$$

then $f \neq g$ since f and g have different codomains.

If $f : \mathbf{Q} \longrightarrow \mathbf{Q}$ and $g : \mathbf{Q} \longrightarrow \mathbf{Q}$ are defined by the rules

$$f(x) = x^2 - 1 \text{ for all } x \in \mathbf{Q}$$

$$g(y) = (y - 1)(y + 1) \text{ for all } y \in \mathbf{Q}$$

then $f = g$.

In certain applications it is important to know the precise codomain of a function, but in many applications the precise codomain is not important, and in such cases I will often omit all mention of the codomain. For example, I might say "For each positive number a, let $J(a) = [0, a]$." and proceed as though I had defined a function. Here you could reasonably take the codomain to be the set of real intervals, or the set of closed intervals, or the set of all subsets of $\mathbf{R}$.

3.48 Definition (Image of f) Let A, B be sets, and let $f : A \longrightarrow B$. The set

$$\{y \in B : \text{ for some } x \in A \ (y = f(x))\}$$

is called the *image of* f, and is denoted by $f(A)$. More generally, if T is any subset of A then we define

$$f(T) = \{y \in B : \text{ for some } x \in T \ (y = f(x))\}.$$

We call $f(T)$ the f-image of T. Clearly, for every subset T of A we have $f(T) \subset B$.

3.49 Examples. If $f : \mathbf{Z} \longrightarrow \mathbf{Z}$ is defined by the rule

$$f(n) = n + 3 \text{ for all } n \in \mathbf{Z}$$

then $f(2) = 5$ so $f(2) \in \mathbf{Z}$,
$f(\{2\}) = \{5\}$ so $f(\{2\}) \subset \mathbf{Z}$,
$f(\mathbf{N}) = \mathbf{Z}_{\geq 3}$.

3.50 Definition (Graph of f) Let A, B be sets, and let $f : A \longrightarrow B$. The *graph of* f is defined to be

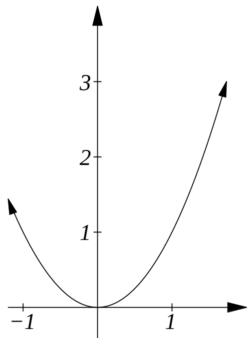
$$\{(x, y) \in A \times B : y = f(x)\}$$

If the domain and codomain of f are subsets of $\mathbf{R}$, then the graph of f can be identified with a subset of the plane.

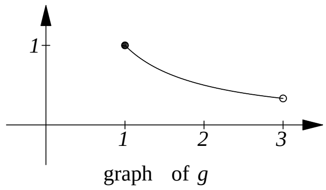
3.51 Examples. Let $f : \mathbf{R} \longrightarrow \mathbf{R}$ be defined by the rule

$$f(x) = x^2 \text{ for all } x \in \mathbf{R}.$$

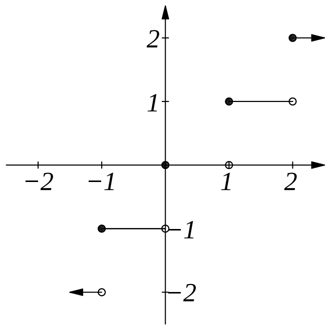
The graph of f is sketched below. The arrowheads on the graph are intended to indicate that the complete graph has not been drawn.



graph of f



graph of g



graph of h

Let $S = \{x \in \mathbf{R} \ : 1 \leq x < 3\}$. Let g be the function from S to $\mathbf{R}$ defined by the rule

$$g(x) = \frac{1}{x} \text{ for all } x \in S.$$

The graph of g is sketched above. The solid dot at $(1, 1)$ indicates that $(1, 1)$ is in the graph. The hollow dot at $(3, 1/3)$ indicates that $(3, 1/3)$ is not in the graph.

Let $h : \mathbf{R} \longrightarrow \mathbf{R}$ be defined by the rule

$$h(x) = \text{ the greatest integer less than or equal to } x.$$

Thus $h(3.14) = 3$ and $h(-3.14) = -4$. The graph of h is sketched above.

The term function (*functio*) was introduced into mathematics by Leibniz [33, page 272 footnote]. During the seventeenth century the ideas of function and curve were usually thought of as being the same, and a curve was often thought of as the path of a moving point. By the eighteenth century the idea of function was associated with "analytic expression". Leonard Euler (1707–1783) gave the following definition:

> A function of a variable quantity is an analytic expression composed in any way whatsoever of the variable quantity and numbers or constant quantities.
>
> Hence every analytic expression, in which all component quantities except the variable z are constants, will be a function of that

> z; Thus $a + 3z$; $az - 4z^2$; $az + b\sqrt{a^2 - z^2}$; c^z; etc. are functions of z[18, page 3].

The use of the notation "$f(x)$" to represent the value of f at x was introduced by Euler in 1734 [29, page 340].

3.52 Exercise. Sketch the graphs of the following functions:

a) $f(x) = (x - 1)^2$ for all $x \in [0, 4]$.

b) $g(x) = (x - 2)^2$ for all $x \in [-1, 3]$.

c) $h(x) = x^2 - 1$ for all $x \in [-2, 2]$.

d) $k(x) = x^2 - 2^2$ for all $x \in [-2, 2]$.

3.4 Summation Notation

Let k and n be integers with $k \leq n$. Let $x_k, x_{k+1}, \ldots x_n$, be real numbers, indexed by the integers from k to n. We define

$$\sum_{i=k}^{n} x_i = x_k + x_{k+1} + \cdots + x_n, \tag{3.53}$$

i.e. $\sum\limits_{i=k}^{n} x_i$ is the sum of all the numbers $x_k, \ldots x_n$. A sum of one number is defined to be that number, so that

$$\sum_{i=k}^{k} x_i = x_k.$$

The "i" in equation (3.53) is a dummy variable, and can be replaced by any symbol that has no meaning assigned to it. Thus

$$\sum_{j=2}^{4} \frac{1}{j} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}.$$

The following properties of the summation notation should be clear from the definition. (Here $c \in \mathbf{R}$, k and n are integers with $k \leq n$ and $x_k, \ldots, x_n$ and $y_k, \ldots y_n$ are real numbers.)

$$\begin{aligned}
\sum_{j=k}^{n} x_j + \sum_{j=k}^{n} y_j &= \sum_{j=k}^{n} (x_j + y_j). \\
c \sum_{j=k}^{n} x_j &= \sum_{j=k}^{n} c x_j. \\
\sum_{j=k}^{n} 1 &= \sum_{j=k}^{n} 1^j = n - k + 1. \\
\left(\sum_{j=k}^{n} x_j \right) + x_{n+1} &= \sum_{j=k}^{n+1} x_j.
\end{aligned}$$

If $x_j \leq y_j$ for all j satisfying $k \leq j \leq n$ then

$$\sum_{j=k}^{n} x_j \leq \sum_{j=k}^{n} y_j.$$

Also

$$\sum_{j=k}^{n} x_j = \sum_{j=k-1}^{n-1} x_{j+1} = \sum_{j=k+1}^{n+1} x_{j-1} = x_k + \cdots + x_n.$$

Using the summation notation, we can rewrite equations (2.9) and (2.23) as

$$\sum_{p=1}^{n} p^2 = \frac{n(n+1)(2n+1)}{6}$$

and

$$\sum_{j=0}^{n-1} r^j = \frac{1 - r^n}{1 - r}.$$

The use of the Greek letter Σ to denote sums was introduced by Euler in 1755[15, page 61]. Euler writes

$$\Sigma x^2 = \frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{6}.$$

Compare this with the notation in Bernoulli's table on page 27. (The apparent difference is due to the fact that for Euler, Σx^2 denotes the sum of x squares, starting with 0^2, whereas for Bernoulli $\int nn$ denotes the sum of n squares starting with 1^2.) The use of the symbol $\int$ (which is a form of S) for sums was introduced by Leibniz. The use of limits on sums was introduced by Augustin Cauchy(1789-1857). Cauchy used the notation $\sum_{m}^{n} {}_r fr$ to denote what we would write as $\sum_{r=m}^{n} f(r)$[15, page 61].

3.54 Exercise. Find the following sums:

a) $\sum_{j=1}^{n}(2j-1)$ for $n = 1, 2, 3, 4$.

b) $\sum_{j=1}^{n} \frac{1}{j(j+1)}$ for $n = 1, 2, 3, 4$.

c) $\sum_{j=1}^{9} \frac{9}{10^j}$.

3.5 Mathematical Induction

The induction principle is a way of formalizing the intuitive idea that if you begin at 1 and start counting "$1, 2, 3, \ldots$", then eventually you will reach any preassigned number (such as for example, 200004).

3.55 Assumption (The Induction Principle) Let k be an integer, and let P be a proposition form over $\mathbf{Z}_{\geq k}$. If

$P(k)$ is true,

and

"for all $n \in \mathbf{Z}_{\geq k}[P(n) \Longrightarrow P(n+1)]$" is true,

then

"for all $n \in \mathbf{Z}_{\geq k}[P(n)]$" is true.

In order to prove "for all $n \in \mathbf{Z}_{\geq k}, P(n)$" by using the induction principle, you should

1. Prove that $P(k)$ is true.
2. Take a generic element n of $\mathbf{Z}_{\geq k}$ and prove $(P(n) \Longrightarrow P(n+1))$.

Recall that the way to prove "$P(n) \Longrightarrow P(n+1)$" is true, is to assume that $P(n)$ is true and show that then $P(n+1)$ must be true.

3.56 Example. We will use the induction principle to do exercise 2.10. For all $n \in \mathbf{Z}_{\geq 1}$, let

$$P(n) = \left[\sum_{p=1}^{n} p^3 = \frac{n^2(n+1)^2}{4} \right].$$

Then $P(1)$ says

$$\sum_{p=1}^{1} p^3 = \frac{(1^2)(1+1)^2}{4}$$

which is true, since both sides of this equation are equal to 1. Now let n be a generic element of $\mathbf{Z}_{\geq 1}$ Then

$$\begin{aligned}
P(n) \quad &\Longleftrightarrow \quad \sum_{p=1}^{n} p^3 = \frac{n^2(n+1)^2}{4} \\
&\Longrightarrow \quad \left(\sum_{p=1}^{n} p^3 \right) + (n+1)^3 = \frac{n^2(n+1)^2}{4} + (n+1)^3 \\
&\Longrightarrow \quad \sum_{p=1}^{n+1} p^3 = \frac{(n+1)^2}{4}(n^2 + 4(n+1)) = \frac{(n+1)^2}{4}(n^2 + 4n + 4) \\
&\Longrightarrow \quad \sum_{p=1}^{n+1} p^3 = \frac{(n+1)^2(n+2)^2}{4} \\
&\Longleftrightarrow \quad P(n+1).
\end{aligned}$$

It follows from the induction principle that $P(n)$ is true for all $n \in \mathbf{Z}_{\geq 1}$, which is what we wanted to prove. |||

3.57 Example. We will show that for all $n \in \mathbf{Z}_{\geq 4}[n! > 2^n]$.
Proof: Define a proposition form P over $\mathbf{Z}_{\geq 4}$ by

$$P(n) = [n! > 2^n].$$

Now $4! = 24 > 16 = 2^4$, so $4! > 2^4$ and thus $P(4)$ is true.

Let n be a generic element of $\mathbf{Z}_{\geq 4}$. Since $n \in \mathbf{Z}_{\geq 4}$, we know that

$$n + 1 \geq 4 + 1 > 2.$$

Hence

$$\begin{aligned} P(n) &\implies (n! > 2^n > 0) \text{ and } (n + 1 > 2 > 0) \\ &\implies (n + 1) \cdot n! > 2 \cdot 2^n \\ &\implies ((n + 1)! > 2^{n+1}) \Longrightarrow P(n + 1). \end{aligned}$$

Hence, for all $n \in \mathbf{Z}_{\geq 4}[P(n) \Longrightarrow P(n + 1)]$. It follows from the induction principle that for all $n \in \mathbf{Z}_{\geq 4}[n! > 2^n]$. ‖

Chapter 4

Analytic Geometry

4.1 Addition of Points

From now on I will denote points in the plane by lower case boldface letters, e.g. $\mathbf{a}, \mathbf{b}, \cdots$. If I specify a point $\mathbf{a}$ and do not explicitly write down its components, you should assume $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, $\cdots, \mathbf{k} = (k_1, k_2)$, etc. The one exception to this rule is that I will always take

$$\mathbf{x} = (x, y).$$

4.1 Definition (Addition of Points) If $\mathbf{a}$ and $\mathbf{b}$ are points in $\mathbf{R}^2$ and $t \in \mathbf{R}$, we define

$$\begin{aligned}
\mathbf{a} + \mathbf{b} &= (a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2) \\
\mathbf{a} - \mathbf{b} &= (a_1, a_2) - (b_1, b_2) = (a_1 - b_1, a_2 - b_2) \\
t\mathbf{a} &= t(a_1, a_2) = (ta_1, ta_2).
\end{aligned}$$

If $t \neq 0$, we will write $\dfrac{\mathbf{a}}{t}$ for $\dfrac{1}{t}\mathbf{a}$; i.e., $\dfrac{\mathbf{a}}{t} = \left(\dfrac{a_1}{t}, \dfrac{b_1}{t}\right)$. We will abbreviate $(-1)\mathbf{a}$ by $-\mathbf{a}$, and we will write $\mathbf{0} = (0, 0)$.

4.2 Theorem. *Let* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *be arbitrary points in* $\mathbf{R}^2$ *and let* s, t *be arbitrary numbers. Then we have:*

Addition is commutative,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

Addition is associative,

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}).$$

We have the following law that resembles the associative law for multiplication:

$$s(t\mathbf{a}) = (st)\mathbf{a}.$$

We have the following distributive laws:

$$(s + t)\mathbf{a} = s\mathbf{a} + t\mathbf{a}, \tag{4.3}$$

$$s(\mathbf{a} + \mathbf{b}) = s\mathbf{a} + s\mathbf{b}. \tag{4.4}$$

Also,

$$1\mathbf{a} = \mathbf{a}, \quad 0\mathbf{a} = \mathbf{0} \ \textit{and}\ \mathbf{a} + (-\mathbf{a}) = \mathbf{0}.$$

All of these properties follow easily from the corresponding properties of real numbers. I will prove the commutative law and one of the distributive laws, and omit the remaining proofs.

Proof of Commutative Law: Let $\mathbf{a}$, $\mathbf{b}$ be points in $\mathbf{R}^2$. By the commutative law for $\mathbf{R}$,

$$a_1 + b_1 = b_1 + a_1 \text{ and } a_2 + b_2 = b_2 + a_2.$$

Hence

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2) = (b_1 + a_1, b_2 + a_2) \\ &= (b_1, b_2) + (a_1, a_2) = \mathbf{b} + \mathbf{a}. \end{aligned}$$

and hence $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.

Proof of (4.3): Let $s, t \in \mathbf{R}$ and let $\mathbf{a} \in \mathbf{R}^2$. By the distributive law for $\mathbf{R}$ we have

$$(s + t)a_1 = sa_1 + ta_1 \text{ and } (s + t)a_2 = sa_2 + ta_2.$$

Hence,

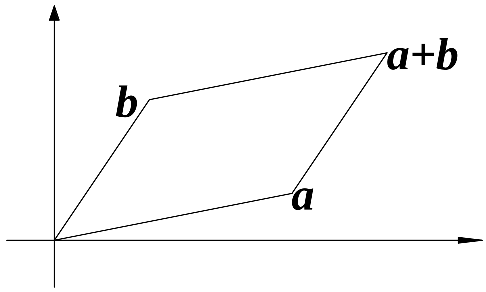
$$\begin{aligned} (s + t)\mathbf{a} &= (s + t)(a_1, a_2) = ((s + t)a_1, (s + t)a_2) = (sa_1 + ta_1, sa_2 + ta_2) \\ &= (sa_1, sa_2) + (ta_1, ta_2) = s(a_1, a_2) + t(a_1, a_2) \\ &= s\mathbf{a} + t\mathbf{a}, \end{aligned}$$

i.e,

$$(s+t)\mathbf{a} = s\mathbf{a} + t\mathbf{a}. \parallel$$

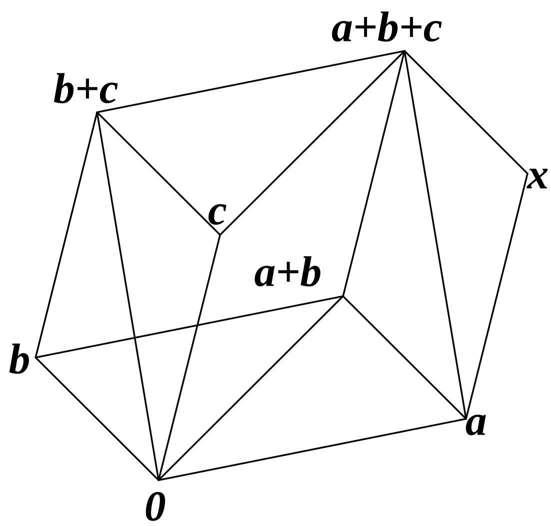
4.5 Notation (Lines in $\mathbf{R}^2$.) If $\mathbf{a}, \mathbf{b}$ are distinct points in $\mathbf{R}^2$, I will denote the (infinite) line through $\mathbf{a}$ and $\mathbf{b}$ by $\mathbf{ab}$, and I will denote the line segment joining $\mathbf{a}$ to $\mathbf{b}$ by $[\mathbf{ab}]$. Hence $[\mathbf{ab}] = [\mathbf{ba}]$.

Remark: Let $\mathbf{a}$, $\mathbf{b}$ be points in $\mathbf{R}^2$ such that $\mathbf{0}$, $\mathbf{a}$ and $\mathbf{b}$ are not all in a straight line. Then $\mathbf{a}+\mathbf{b}$ is the vertex opposite $\mathbf{0}$ in the parallelogram whose other three vertices are $\mathbf{b}$, $\mathbf{0}$ and $\mathbf{a}$.



Proof: In this proof I will suppose $a_1 \neq 0$ and $b_1 \neq 0$, so that neither of $\mathbf{0a}, \mathbf{0b}$ is a vertical line. (I leave the other cases to you.) The slope of line $\mathbf{0a}$ is $\frac{a_2 - 0}{a_1 - 0} = \frac{a_2}{a_1}$, and the slope of $\mathbf{b}(\mathbf{a}+\mathbf{b})$ is $\frac{(a_2+b_2)-b_2}{(a_1+b_1)-b_1} = \frac{a_2}{a_1}$. Thus the lines $\mathbf{0a}$ and $\mathbf{b}(\mathbf{a}+\mathbf{b})$ are parallel.

The slope of line $\mathbf{0b}$ is $\frac{b_2 - 0}{b_1 - 0} = \frac{b_2}{b_1}$, and the slope of $\mathbf{a}(\mathbf{a}+\mathbf{b})$ is $\frac{(a_2+b_2)-a_2}{(a_1+b_1)-a_1} = \frac{b_2}{b_1}$. Thus the lines $\mathbf{0b}$ and $\mathbf{a}(\mathbf{a}+\mathbf{b})$ are parallel. It follows that the figure $\mathbf{0a}(\mathbf{a}+\mathbf{b})\mathbf{b}$ is a parallelogram, i.e., $\mathbf{a}+\mathbf{b}$ is the fourth vertex of a parallelogram having $\mathbf{0}$, $\mathbf{a}$, and $\mathbf{b}$ as its other vertices. $\parallel$



4.6 Example. In the figure you should be able to see the parallelograms defining $\mathbf{a}+\mathbf{b}$, $(\mathbf{a}+\mathbf{b})+\mathbf{c}$, $\mathbf{b}+\mathbf{c}$ and $\mathbf{a}+(\mathbf{b}+\mathbf{c})$. Also you should be able to see geometrically that $(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$. What is the point marked $\mathbf{x}$ in the figure?

4.7 Exercise. In figure a), $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{e}$, and $\mathbf{f}$ are the vertices of a regular hexagon centered at $\mathbf{0}$. Sketch the points $\mathbf{a}+\mathbf{b}$, $(\mathbf{a}+\mathbf{b})+\mathbf{c}$, $(\mathbf{a}+\mathbf{b}+\mathbf{c})+\mathbf{d}$, $(\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d})+\mathbf{e}$, and $(\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}+\mathbf{e})+\mathbf{f}$ as accurately as you can.

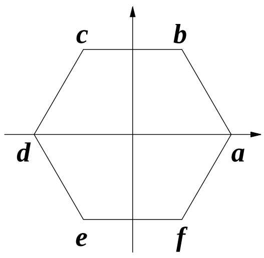


figure a

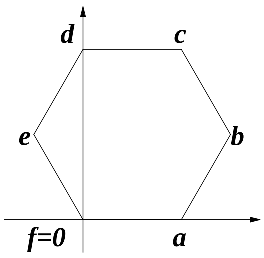


figure b

In figure b), $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{e}$ and $\mathbf{f}$ are the vertices of a regular hexagon with $\mathbf{f}=\mathbf{0}$. Sketch the points $\mathbf{a}+\mathbf{b}$, $(\mathbf{a}+\mathbf{b})+\mathbf{c}$, $(\mathbf{a}+\mathbf{b}+\mathbf{c})+\mathbf{d}$, and $(\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d})+\mathbf{e}$ as accurately as you can. (This problem should be done geometrically. Do not calculate the coordinates of any of these points.)

4.8 Example (Line segment) We will now give an analytical description for a non-vertical line segment $[\mathbf{ab}]$, $(a_1 \neq b_1)$. Suppose first that $a_1 < b_1$. The equation for the line through $\mathbf{a}$ and $\mathbf{b}$ is

$$y = a_2 + \frac{b_2 - a_2}{b_1 - a_1}(x - a_1).$$

Hence a point (x, y) is in $[\mathbf{ab}]$ if and only if there is a number $x \in [a_1, b_1]$ such that

$$\begin{aligned}
(x, y) &= \Big(x, a_2 + \frac{b_2 - a_2}{b_1 - a_1}(x - a_1)\Big) \\
&= \Big(a_1 + (x - a_1), a_2 + \frac{b_2 - a_2}{b_1 - a_1}(x - a_1)\Big) \\
&= (a_1, a_2) + (x - a_1)(1, \frac{b_2 - a_2}{b_1 - a_1}) \\
&= \mathbf{a} + \frac{x - a_1}{b_1 - a_1}(b_1 - a_1, b_2 - a_2) \\
&= \mathbf{a} + \frac{x - a_1}{b_1 - a_1}(\mathbf{b} - \mathbf{a}).
\end{aligned}$$

Now

$$\begin{aligned}
x \in [a_1, b_1] &\iff a_1 \leq x \leq b_1 \\
&\iff 0 \leq x - a_1 \leq b_1 - a_1 \\
&\iff 0 \leq \frac{x - a_1}{b_1 - a_1} \leq 1.
\end{aligned}$$

Thus

$$[\mathbf{ab}] = \{\mathbf{a} + t(\mathbf{b} - \mathbf{a}) : 0 \leq t \leq 1\}.$$

If $b_1 < a_1$ then

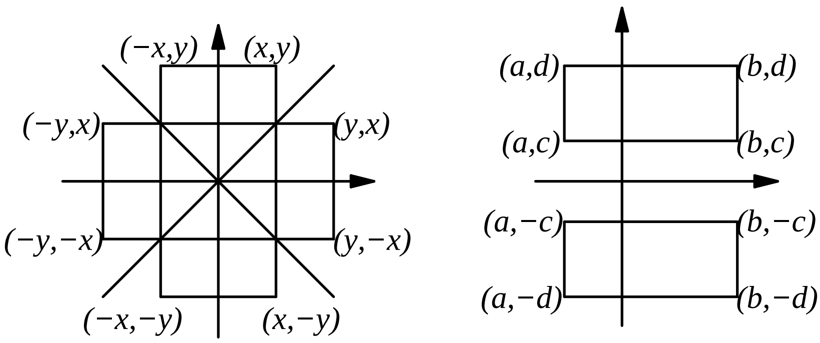
$$\begin{aligned}
[\mathbf{ab}] &= [\mathbf{ba}] = \{\mathbf{b} + t(\mathbf{a} - \mathbf{b}) : 0 \leq t \leq 1\} \\
&= \{\mathbf{a} + (1 - t)(\mathbf{b} - \mathbf{a}) : 0 \leq t \leq 1\}.
\end{aligned}$$

Now as t runs through all values in $[0, 1]$, we see that $1 - t$ also takes on all values in $[0, 1]$ so we get the same description for $[\mathbf{ab}]$ when $b_1 < a_1$ as we do when $a_1 < b_1$. Note that this description is exactly what you would expect from the pictures, and that it also works for vertical segments.

4.2 Reflections, Rotations and Translations

4.9 Definition (Reflections and Rotations.) We now define a family of functions from $\mathbf{R}^2$ to $\mathbf{R}^2$. If $(x, y) \in \mathbf{R}^2$, we define

$$
\begin{aligned}
I(x,y) &= (x,y) && \text{(Identity function.)} && (4.10)\\
H(x,y) &= (x,-y) && \text{(Reflection of } (x,y) \text{ about the horizontal axis.)}\\
V(x,y) &= (-x,y) && \text{(Reflection of } (x,y) \text{ about the vertical axis.)}\\
D_+(x,y) &= (y,x) && \text{(Reflection of } (x,y) \text{ about the line } y=x.)\\
D_-(x,y) &= (-y,-x) && \text{(Reflection of } (x,y) \text{ about the line } y=-x.)\\
R_{\pi/2}(x,y) &= (y,-x) && \text{(Clockwise rotation of } (x,y) \text{ by } \tfrac{\pi}{2}.)\\
R_{-\frac{\pi}{2}}(x,y) &= (-y,x) && \text{(Counter-clockwise rotation of } (x,y) \text{ by } \tfrac{\pi}{2}.)\\
R_\pi(x,y) &= (-x,-y) && \text{Rotation by } \pi. && (4.11)
\end{aligned}
$$



Each of the eight functions just defined carries every box to another box with the same area. You should be able to see from the picture that

$$H\left(B(a, b\colon c, d)\right) = B(a, b\colon -d, -c).$$

We can see this analytically as follows:

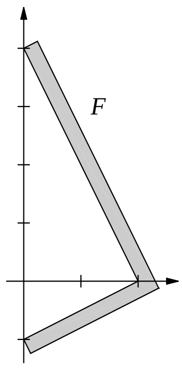
$$
\begin{aligned}
(x,y) \in B(a,b\colon c,d) &\iff a \le x \le b \text{ and } c \le y \le d\\
&\iff a \le x \le b \text{ and } -d \le -y \le -c\\
&\iff (x,-y) \in B(a,b\colon -d,-c)\\
&\iff H(x,y) \in B(a,b\colon -d,-c).
\end{aligned}
$$

I will usually omit the analytic justification in cases like this.

Each of the eight functions described in definition 4.9 carries the square $B(-1, 1{:}\, -1, 1)$ to itself.

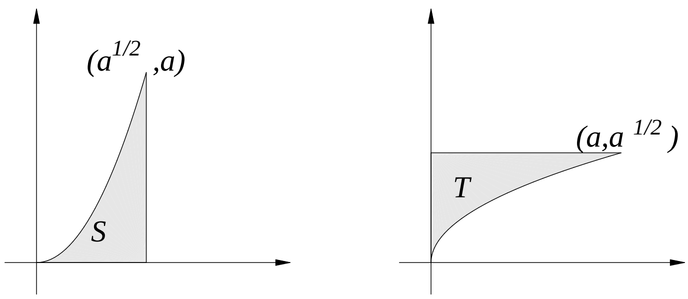
4.12 Definition (Symmetry of the square.) The eight functions defined in equations (4.10)-(4.11) are called *symmetries of the square*.

4.13 Exercise. Let F be the set shown in the figure. On one set of axes draw the sets $F, R_{\pi/2}(F), R_{-\frac{\pi}{2}}(F)$ and $R_\pi(F)$ (label the four sets). On another set of axes draw and label the sets $V(F), H(F), D_+(F)$ and $D_-(F)$.



4.14 Example. Let $a \in \mathbf{R}^+$ and let

$$\begin{aligned} S &= \{(x, y){:}\, 0 \le x \le \sqrt{a} \text{ and } 0 \le y \le x^2\} \\ T &= \{(x, y){:}\, 0 \le x \le a \text{ and } \sqrt{x} \le y \le a\}. \end{aligned}$$



From the picture it is clear that $D_+(S) = T$. An analytic proof of this result is as follows:

$$(x, y) \in S \iff 0 \le x \le \sqrt{a} \text{ and } 0 \le y \le x^2 \tag{4.15}$$

$$\begin{aligned}
&\Longrightarrow \quad 0 \le y \le x^2 \le (\sqrt{a})^2 \text{ and } 0 \le \sqrt{y} \le x \le \sqrt{a} \\
&\Longrightarrow \quad 0 \le y \le a \text{ and } \sqrt{y} \le x \le \sqrt{a} \qquad (4.16) \\
&\Longleftrightarrow \quad (y, x) \in T \\
&\Longleftrightarrow \quad D_+(x, y) \in T.
\end{aligned}$$

To show that $D_+(x,y) \in T \Longrightarrow (x,y) \in S$, I need to show that (4.16) implies (4.15). This follows because

$$0 \le y \le a \text{ and } \sqrt{y} \le x \le \sqrt{a} \Longrightarrow 0 \le x \le \sqrt{a} \text{ and } 0 \le y = (\sqrt{y})^2 \le x^2.$$

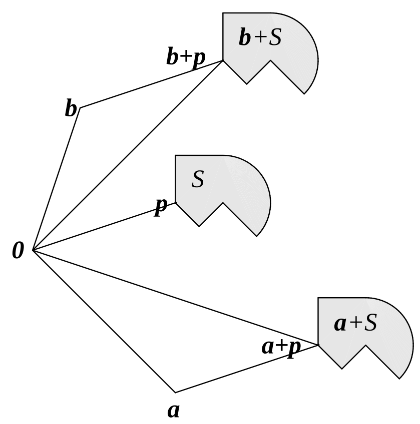
In exercise 2.18 you assumed that S and T have the same area. In general we will assume that if S is a set and F is a symmetry of the square, then S and $F(S)$ have the same area. (Cf. Assumption 5.11.)

4.17 Definition (Translate of a set.) Let S be a set in $\mathbf{R}^2$ and let $\mathbf{a} \in \mathbf{R}^2$. We define the set $\mathbf{a} + S$ by

$$\mathbf{a} + S = \{\mathbf{a} + \mathbf{s} \colon \mathbf{s} \in S\}.$$

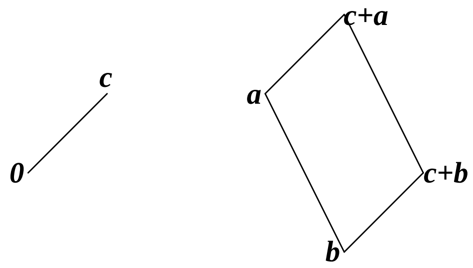
Sets of the form $\mathbf{a} + S$ will be called *translates of S*.

4.18 Example. The pictures below show some examples of translates. Intuitively each translate of S has the same shape as S and each translate of S has the same area as S.



4.19 Example (Translates of line segments.) Let $\mathbf{a}, \mathbf{b} \in \mathbf{R}^2$. If $\mathbf{c} \in \mathbf{R}^2$, then

$$\begin{aligned}\mathbf{c} + [\mathbf{ab}] &= \mathbf{c} + \{\mathbf{a} + t(\mathbf{b} - \mathbf{a}) : 0 \leq t \leq 1\} \\ &= \{\mathbf{c} + \mathbf{a} + t(\mathbf{b} - \mathbf{a}) \colon 0 \leq t \leq 1\} \\ &= \{\mathbf{c} + \mathbf{a} + t\left((\mathbf{c} + \mathbf{b}) - (\mathbf{c} + \mathbf{a})\right) \colon 0 \leq t \leq 1\} \\ &= [(\mathbf{c} + \mathbf{a})(\mathbf{c} + \mathbf{b})].\end{aligned}$$



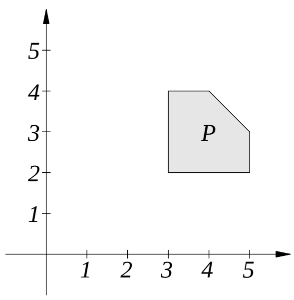
In particular $-\mathbf{a} + [\mathbf{a}, \mathbf{b}] = [\mathbf{0}, \mathbf{b} - \mathbf{a}]$, so any segment can be translated to a segment with $\mathbf{0}$ as an endpoint.

4.20 Exercise. Let a, b, c, d, r, s be real numbers with $a \leq b$ and $c \leq d$. Show that

$$(r, s) + B(a, b \colon c, d) = B(?, ?; ?, ?)$$

if the four question marks are replaced by suitable expressions. Include some explanation for your answer.

4.21 Exercise. Let P be the set shown in the figure below.



a) Sketch the sets $(-2, -2) + P$ and $(4, 1) + P$.

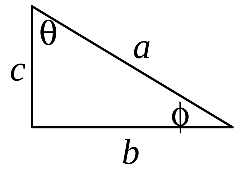
b) Sketch the sets $R_{\frac{\pi}{2}}((1, 1) + P)$ and $(1, 1) + R_{\frac{\pi}{2}}(P)$, where $R_{\frac{\pi}{2}}$ is defined as in definition 4.9

4.3 The Pythagorean Theorem and Distance.

Even though you are probably familiar with the Pythagorean theorem, the result is so important and non-obvious that I am including a proof of it.

4.22 Theorem (Pythagorean Theorem.) *In any right triangle, the square on the hypotenuse is equal to the sum of the squares on the two legs.*

Proof: Consider a right triangle T whose legs have length b and c, and whose hypotenuse has length a, and whose angles are ϕ and θ as shown in the figure.



We have $\phi + \theta = 90°$ since T is a right triangle.

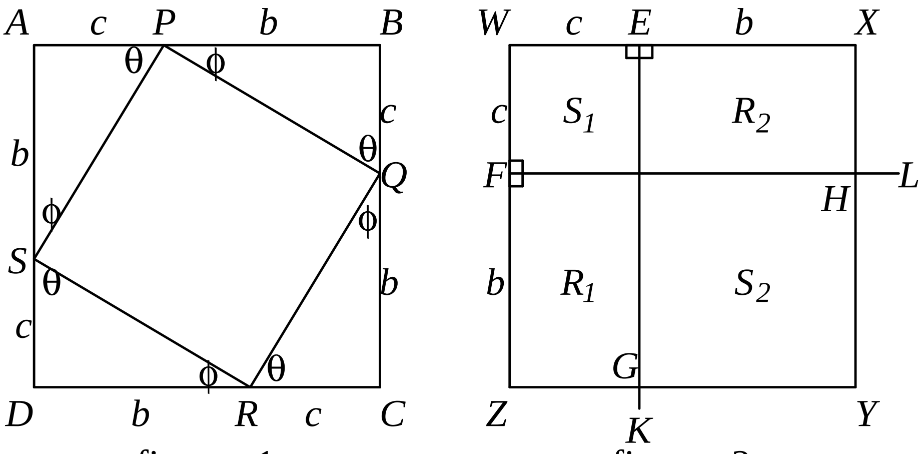


figure 1 figure 2

Construct a square $ABCD$ with sides of length $b+c$, and find points P, Q, R, S dividing the sides of $ABCD$ into pieces of sizes b and c as shown in figure 1. Draw the lines PQ, QR, RS, and SP, thus creating four triangles congruent to T (i.e., four right triangles with legs of length b and c). Each angle of $PQRS$

is $180° - (\phi + \theta) = 180° - 90° = 90°$ so $PQRS$ is a square of side a. The four triangles in figure 1 each have area $\frac{1}{2}bc$, so

$$\text{area}(ABCD) - 4 \cdot \text{area}(T) = a^2 \tag{4.23}$$

or

$$(b + c)^2 - 2bc = a^2$$

and hence

$$b^2 + c^2 = a^2 \; \| \tag{4.24}$$

The proof just given uses a combination of algebra and geometry. I will now give a second proof that is completely geometrical.

Construct a second square $WXYZ$ with sides of length $b + c$, and mark off segments WE and WF of length c as shown in figure 2. Then draw EK perpendicular to WX and let EK intersect ZY at G, and draw FL perpendicular to WZ and let FL intersect XY at H. Then EGZ is a right angle, since the other angles of the quadrilateral $WEGZ$ are right angles. Similarly angle FHX is a right angle. Thus $WEGZ$ is a rectangle so $ZG = c$ and similarly $WFHX$ is a rectangle and $XH = c$. Moreover EG and FH are perpendicular since $EG \| WZ$ and $FH \| WX$. Thus the region labeled S_1 is a square with side c and the region labeled S_2 is a square with side b.

In figure 2 we have $\text{area}(R_1) = \text{area}(R_2) = 2\text{area}(T)$, and hence

$$\text{area}(WXYZ) - 4 \cdot \text{area}(T) = b^2 + c^2. \tag{4.25}$$

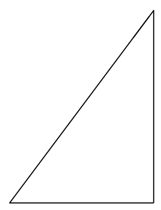
We have $\text{area}(ABCD) = \text{area}(WXYZ)$ since $ABCD$ and $WXYZ$ are both squares with side $b + c$. Hence from equations (4.23) and (4.25) we see that

$$a^2 = b^2 + c^2. \; \|$$

Although the theorem we just proved is named for Pythagoras (fl. 530–510 B.C) , it was probably known much earlier. There is evidence that it was known to the Babylonians circa 1000 BC[27, pp 118-121]. Legend has it that

> Emperor Yǔ[circa 21st century B.C.] quells floods, he deepens rivers and streams, observes the shape of mountains and valleys, surveys the high and low places, relieves the greatest calamities

> and saves the people from danger. He leads the floods east into the sea and ensures no flooding or drowning. This is made possible because of the Gōugŭ theorem ...[47, page 29].



Gōugŭ shape

"Gōugŭ" is the shape shown in the figure, and the Gōugŭ theorem is our Pythagorean theorem. The prose style here is similar to that of current day mathematicians trying to get congress to allocate funds for the support of mathematics.

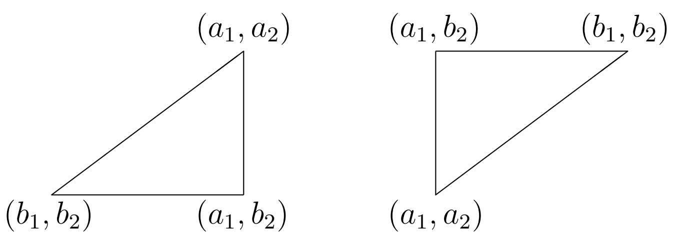
Katyayana(c. 600 BC or 500BC??) stated the general theorem:

> The rope [stretched along the length] of the diagonal of a rectangle makes an [area] which the vertical and horizontal sides make together.[27, page 229]

4.26 Theorem (Distance formula.) *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are points in* $\mathbf{R}^2$ *then the distance from* $\mathbf{a}$ *to* $\mathbf{b}$ *is*

$$d(\mathbf{a},\mathbf{b}) = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}.$$

Proof: Draw the vertical line through $\mathbf{a}$ and the horizontal line through $\mathbf{b}$. These lines intersect at the point $\mathbf{p} = (a_1, b_2)$. The length of $[\mathbf{ap}]$ is $|a_2 - b_2|$ and the length of $[\mathbf{pb}]$ is $|a_1 - b_1|$ and $[\mathbf{ab}]$ is the hypotenuse of a right angle with legs $[\mathbf{ap}]$ and $[\mathbf{pb}]$.



By the Pythagorean theorem,

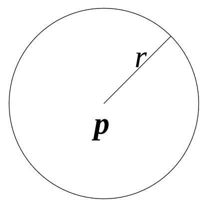
$$(\text{length}([\mathbf{ab}]))^2 = (a_1 - b_1)^2 + (a_2 - b_2)^2$$

so $\text{length}([\mathbf{a}, \mathbf{b}]) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$. |||

4.27 Notation ($d(\mathbf{a}, \mathbf{b})$, distance$(\mathbf{a}, \mathbf{b})$) If $\mathbf{a}$ and $\mathbf{b}$ are points in $\mathbf{R}^2$, I will denote the distance from $\mathbf{a}$ to $\mathbf{b}$ by either distance$(\mathbf{a}, \mathbf{b})$ or by $d(\mathbf{a}, \mathbf{b})$.

4.28 Definition (Circle.) Let $\mathbf{p} = (a, b)$ be a point in $\mathbf{R}^2$, and let $r \in \mathbf{R}^+$. The *circle with center* $\mathbf{p}$ *and radius* r is defined to be

$$\begin{aligned} C(\mathbf{p}, r) &= \{(x, y) \in \mathbf{R}^2 \colon d((x, y), (a, b)) = r\} \\ &= \{(x, y) \in \mathbf{R}^2 \colon \sqrt{(x - a)^2 + (y - b)^2} = r\} \\ &= \{(x, y) \in \mathbf{R}^2 \colon (x - a)^2 + (y - b)^2 = r^2\}. \end{aligned}$$



C(p,r)

The equation

$$(x - a)^2 + (y - b)^2 = r^2$$

is called the *equation of the circle* $C(\mathbf{p}, r)$. The circle $C((0, 0), 1)$ is called the *unit circle.*

We will now review the method for solving quadratic equations.

4.29 Theorem (Quadratic formula.) *Let A, B, and C be real numbers with $A \neq 0$.*

If $B^2 - 4AC < 0$, then the equation $Ax^2 + Bx + C = 0$ has no solutions in $\mathbf{R}$.

If $B^2 - 4AC \geq 0$, then the set of solutions of the equation $Ax^2 + Bx + C = 0$ is

$$\left\{ \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \right\}. \tag{4.30}$$

The set (4.30) contains one or two elements, depending on whether $B^2 - 4AC$ is zero or positive.)

Proof: Let A, B, C be real numbers with $A \neq 0$. Let $x \in \mathbf{R}$. Then

$$\begin{aligned} Ax^2 + Bx + C = 0 \quad &\iff \quad A\Big(x^2 + \frac{Bx}{A}\Big) = -C \\ &\iff \quad A\Big(x^2 + \frac{Bx}{A} + \frac{B^2}{4A^2}\Big) = -C + \frac{AB^2}{4A^2} = -C + \frac{B^2}{4A} \\ &\iff \quad A\Big(x + \frac{B}{2A}\Big)^2 = \frac{-4AC + B^2}{4A} \\ &\iff \quad \Big(x + \frac{B}{2A}\Big)^2 = \frac{B^2 - 4AC}{4A^2} \end{aligned}$$

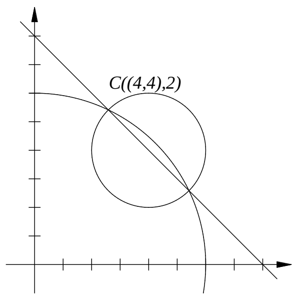
Hence $Ax^2+Bx+C = 0$ has no solutions unless $B^2-4AC \geq 0$. If $B^2-4AC \geq 0$, then the solutions are given by

$$\Big(x + \frac{B}{2A}\Big) = \frac{\pm\sqrt{B^2 - 4AC}}{2A}$$

i.e.,

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}. \quad \|$$

4.31 Example. Describe the set $C((0,0),6) \cap C((4,4),2)$.



The sketch suggests that this set will consist of two points in the first quadrant. Let (x, y) be a point in the intersection. Then

$$x^2 + y^2 = 36 \tag{4.32}$$

and

$$(x-4)^2 + (y-4)^2 = 4, \text{ i.e. } x^2 + y^2 - 8x - 8y + 28 = 0. \tag{4.33}$$

It follows that $36 - 8x - 8y + 28 = 0$, or $8x + 8y - 64 = 0$ or

$$y = 8 - x. \tag{4.34}$$

(The line whose equation is $y = 8 - x$ is shown in the figure. We've proved that the intersection is a subset of this line.) Replace y by $8 - x$ in equation (4.33) to obtain

$$x^2 + (8 - x)^2 - 8x - 8(8 - x) + 28 = 0$$

i.e.,

$$x^2 + 64 - 16x + x^2 - 8x - 64 + 8x + 28 = 0$$

i.e.,

$$2x^2 - 16x + 28 = 0$$

i.e.,

$$x^2 - 8x + 14 = 0.$$

By the quadratic formula, it follows that

$$x = \frac{8 \pm \sqrt{64 - 56}}{2} = \frac{8 \pm 2\sqrt{2}}{2} = 4 \pm \sqrt{2}.$$

By equation (4.34)

$$y = 8 - x = 4 \mp \sqrt{2}.$$

We have shown that if $(x, y) \in C((0, 0), 6) \cap C((4, 4), 2)$, then $(x, y) \in \{(4 + \sqrt{2}, 4 - \sqrt{2}), (4 - \sqrt{2}, 4 + \sqrt{2})\}$. It is easy to verify that each of the two calculated points satisfies both equations (4.32) and (4.33) so

$$C((0, 0), 6) \cap C((4, 4), 2) = \{(4 + \sqrt{2}, 4 - \sqrt{2}), (4 - \sqrt{2}, 4 + \sqrt{2})\}.$$

Chapter 5

Area

In chapter 2 we calculated the area of the set

$$\{(x, y) \in \mathbf{R}^2 \colon 0 \leq x \leq a \text{ and } 0 \leq y \leq x^2\}$$

where $a \geq 0$, and of the set

$$\{(x, y) \in \mathbf{R}^2 \colon 1 \leq x \leq b \text{ and } 0 \leq y \leq x^{-2}\}$$

where $b > 1$.

The technique that was used for making these calculations can be used to find the areas of many other subsets of $\mathbf{R}^2$. The general procedure we will use for finding the area of a set S will be to find two sequences of polygons $\{I_n\}$ and $\{O_n\}$ such that

$$I_n \subset S \subset O_n \text{ for all } n \in \mathbf{Z}^+.$$

We will then have

$$\text{area}(I_n) \leq \text{area}(S) \leq \text{area}(O_n) \text{ for all } n \in \mathbf{Z}^+. \tag{5.1}$$

We will construct the polygons I_n and O_n so that $\text{area}(O_n) - \text{area}(I_n)$ is arbitrarily small when n large enough, and we will see that then there is a unique number A such that

$$\text{area}(I_n) \leq A \leq \text{area}(O_n) \text{ for all } n \in \mathbf{Z}^+. \tag{5.2}$$

We will take A to be the area of S.

5.1 Basic Assumptions about Area

5.3 Definition (Bounded Sets.) A subset S of $\mathbf{R}^2$ is *bounded* if $S \subset B(a, b\colon c, d)$ for some box $B(a, b\colon c, d)$. A subset of $\mathbf{R}^2$ that is not bounded is said to be *unbounded*.

It is clear that every subset of a bounded set is bounded. It is not difficult to show that if B is a bounded set then $\mathbf{a} + B$ is bounded for every $\mathbf{a} \in \mathbf{R}^2$, and $S(B)$ is bounded for every symmetry of the square, S.

5.4 Example. The set

$$\{(n, \frac{1}{n}) : n \in \mathbf{Z}^+\}$$

is an unbounded subset of $\mathbf{R}^2$. I cannot draw a picture of an unbounded set, because the sheet of paper on which I make my drawing will represent a box containing any picture I draw.

5.5 Definition (Bounded Function.) Let S be a set. A function $f : S \to \mathbf{R}$ is called a *bounded function* if there is a number M such that $|f(x)| \leq M$ for all $x \in S$. It is clear that if f is a bounded function on an interval $[a, b]$, then graph(f) is a bounded subset of $\mathbf{R}^2$, since graph$(f) \subset B(a, b : -M, M)$. If f is bounded on S, then any number M satisfying

$$|f(x)| \leq M \text{ for all } x \in S$$

is called a *bound for f* on S.

We are now ready to state our official assumptions about area. At this point you should officially forget everything you know about area. Unofficially, however, you remember everything you know so that you can evaluate whether the theorems we prove are reasonable. Our aim is not simply to calculate areas, but to see how our calculations are justified by our assumptions.

We will assume that there is a function α from the set of bounded subsets of $\mathbf{R}^2$ to the real numbers that satisfies the conditions of positivity, additivity, normalization, translation invariance and symmetry invariance described below. Any function α that satisfies these conditions will be called an *area function*.

5.6 Assumption (Positivity of area.)

$$\alpha(S) \geq 0 \text{ for every bounded subset } S \text{ of } \mathbf{R}^2.$$

5.7 Definition (Disjoint sets.) We say two sets S, T are *disjoint* if and only if $S \cap T = \emptyset$.

5.8 Assumption (Additivity of area.) If S, T are disjoint bounded subsets of $\mathbf{R}^2$, then

$$\alpha(S \cup T) = \alpha(S) + \alpha(T).$$

5.9 Assumption (Normalization property of area.) For every box $B(a, b\colon c, d)$ we have

$$\alpha\Big(B(a, b\colon c, d)\Big) = (b - a)(d - c),$$

i.e., the area of a box is the product of the length and the width of the box.

5.10 Assumption (Translation invariance of area.) Let S be a bounded set in $\mathbf{R}^2$, and let $\mathbf{a} \in \mathbf{R}^2$, then

$$\alpha(S) = \alpha(\mathbf{a} + S).$$

5.11 Assumption (Invariance under symmetry.) Let S be a bounded subset of $\mathbf{R}^2$. Then if F is any symmetry of the square

$$\alpha(F(S)) = \alpha(S).$$

(See definition 4.12 for the definition of symmetry of the square.)

Remark: I would like to replace the assumptions 5.10 and 5.11 by the single stronger assumption:

> If A and B are bounded subsets of $\mathbf{R}^2$, and A is congruent to B, then $\alpha(A) = \alpha(B)$.

However the problem of defining what *congruent* means is rather complicated, and I do not want to consider it at this point.

5.12 Entertainment (Congruence problem.) Formulate a definition of what it means for two subsets of $\mathbf{R}^2$ to be congruent.

5.13 Example. Let

$$\begin{aligned} S &= B(0,1:0,1) \cap \{(x,y) \in \mathbf{R}^2 : x \in \mathbf{Q}\} \\ T &= B(0,1:0,1) \cap \{(x,y) \in \mathbf{R}^2 : x \notin \mathbf{Q}\}. \end{aligned}$$

I do not know how to make any reasonable drawing of S or T. Any picture I draw of S would look just like a picture of T, even though the two sets are disjoint. By additivity and the normalization property for area

$$\alpha(S) + \alpha(T) = \alpha(S \cup T) = \alpha(B(0,1:0,1)) = 1.$$

Since areas are non-negative, it follows that

$$0 \le \alpha(S) \le 1 \text{ and } 0 \le \alpha(T) \le 1.$$

The problem of calculating $\alpha(S)$ exactly cannot be answered on the basis of the assumptions we have made.

Remarks: The assumptions we have just made are supposed to be intuitively plausible. When we choose to make a particular set of assumptions, we hope that the assumptions are consistent, i.e., that no contradictions follow from them. If we were to add a new assumption:

The area of a circle with radius 1 is 3.14159,

then we would have an inconsistent set of assumptions, because it follows from the assumptions we have already made that the area of a circle of radius 1 is greater than 3.141592.

In 1923 Stefan Banach(1892–1945) [5] showed that area functions exist, i.e., that the assumptions we have made about area are consistent. Unfortunately Banach showed that there is more than one area function, and different area functions assign different values to the set S described in the previous example.

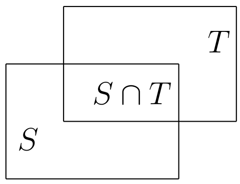
A remarkable result of Felix Hausdorff(1868–1942) [24, pp469–472] shows that the analogous assumptions for volume in three dimensional space (if we include the assumption that any two congruent sets in 3 dimensional space $\mathbf{R}^3$ have the same volume) are inconsistent. If one wants to discuss volume in $\mathbf{R}^3$ then one cannot consider volumes of arbitrary sets. One must considerably restrict the class of sets that have volumes. A discussion of Hausdorff's result can be found in [20].

5.2 Further Assumptions About Area

In this section we will introduce some more assumptions about area. The assumptions in this section can actually be proved on the basis of the basic assumptions we have already made, and in fact the proofs are easy (the proofs are given in appendix B). The reason I have made assumptions out of them is that they are as intuitively plausible as the assumptions I have already made, and I do not have time to do everything I want to do. I am omitting the proofs with regret because I agree with Aristotle that

> It is manifest that it is far better to make the principles finite in number. Nay, they should be the fewest possible provided they enable all the same results to be proved. This is what mathematicians insist upon; for they take as principles things finite either in kind or in number.[25, page 178]

5.14 Assumption (Addition rule for area.)



For any bounded sets S and T in $\mathbf{R}^2$

$$\alpha(S \cup T) = \alpha(S) + \alpha(T) - \alpha(S \cap T). \tag{5.15}$$

and consequently

$$\alpha(S \cup T) \leq \alpha(S) + \alpha(T).$$

5.16 Assumption (Subadditivity of area.) Let $n \in \mathbf{Z}_{\geq 1}$, and let A_1, A_2, $\cdots$, A_n be bounded sets in $\mathbf{R}^2$. Then

$$\alpha(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \alpha(A_i). \tag{5.17}$$

5.18 Assumption (Monotonicity of area.) Let S, T be bounded sets such that $S \subset T$. Then $\alpha(S) \leq \alpha(T)$.

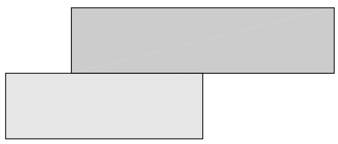
5.19 Definition (Zero-area set.) We will call a set with zero area a *zero-area set.*

From the normalization property it follows that every horizontal or vertical segment has area equal to 0. Thus every horizontal or vertical segment is a zero-area set.

5.20 Corollary (to assumption 5.18.) [1] *Every subset of a zero-area set is a zero-area set. In particular the empty set is a zero-area set.*

5.21 Corollary (to assumption 5.16.) *The union of a finite number of zero-area sets is a zero-area set.*

5.22 Definition (Almost disjoint.) We will say that two bounded subsets S, T of $\mathbf{R}^2$ are *almost disjoint* if $S \cap T$ is a zero-area set.



Almost disjoint sets

5.23 Examples. If a, b, c are real numbers with $a < b < c$, then since

$$B(a, b{:}\, p, q) \cap B(b, c{:}\, s, t) \subset B(b, b{:}\, p, q),$$

the boxes $B(a, b{:}\, p, q)$ and $B(b, c{:}\, s, t)$ are almost disjoint.

Any zero-area set is almost disjoint from every set – including itself.

5.24 Assumption (Additivity for almost disjoint sets.) Let $\{R_1, \cdots, R_n\}$ be a finite set of bounded sets such that R_i and R_j are almost disjoint whenever $i \neq j$. Then

$$\alpha(\bigcup_{i=1}^{n} R_i) = \sum_{i=1}^{n} \alpha(R_i). \tag{5.25}$$

[1] Usually a corollary is attached to a *theorem* and not to an *assumption*. A corollary is a statement that follows immediately from a theorem without a proof. By etymology, it is a "small gift".

5.26 Notation (Area functions α or area**)** Any real valued function α, whose domain is the family of bounded subsets of $\mathbf{R}^2$, and which satisfies all of the assumptions listed in sections 5.1 and 5.2 will be called an *area function.* In these notes I will use the names "α" and "area" to denote area function. Thus

$$\alpha(B(a,b:c,d)) = \text{area}(B(a,b:c,d)) = (b-a)(d-c).$$

5.3 Monotonic Functions

5.27 Definition (Partition.) Let a, b be real numbers with $a \le b$. A *partition* P of the interval $[a,b]$ is a finite sequence of points

$$P = \{x_0, x_1, \cdots, x_n\}$$

with $a = x_0 \le x_1 \le x_2 \cdots \le x_n = b$. The intervals $[x_0, x_1], [x_1, x_2], \cdots,$ $[x_{n-1}, x_n]$ are called the *subintervals of the partition* P, and $[x_{j-1}, x_j]$ is the j^{th} *subinterval of* P for $1 \le j \le n$. The largest of the numbers $x_j - x_{j-1}$ is called the *mesh of the partition* P, and is denoted by $\mu(P)$. The partition

$$\{a,\ a + \frac{(b-a)}{n},\ a + \frac{2(b-a)}{n}, \cdots, a + \frac{n(b-a)}{n} = b\}$$

is called the *regular partition of* $[a,b]$ *into* n *equal subintervals.*

5.28 Example. Let

$$P = \{0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\}$$

Then P is a partition of $[0,1]$ into 5 subintervals and $\mu(P) = 1 - \frac{1}{2} = \frac{1}{2}$.

The regular partition of $[1,2]$ into 5 subintervals is $\{1, \frac{6}{5}, \frac{7}{5}, \frac{8}{5}, \frac{9}{5}, 2\}$.

If Q_n is the regular partition of $[a,b]$ into n equal subintervals, then $\mu(Q_n) = \frac{b-a}{n}$.

5.29 Exercise. Find a partition P of $[0,1]$ into five subintervals, such that $\mu(P) = \frac{4}{5}$, or explain why no such partition exists.

5.30 Exercise. Find a partition Q of $[0,1]$ into five subintervals, such that $\mu(Q) = \frac{1}{10}$, or explain why no such partition exists.

5.31 Definition (Monotonic function.) Let J be an interval in $\mathbf{R}$, and let $f\colon J \to \mathbf{R}$ be a function. We say that f is *increasing on J* if

$$\text{for all } x, y \text{ in } J\Big[(x \leq y) \implies (f(x) \leq f(y))\Big] \tag{5.32}$$

and we say that f is *decreasing on J* if

$$\text{for all } x, y \text{ in } J\Big[(x \leq y) \implies (f(x) \geq f(y))\Big].$$

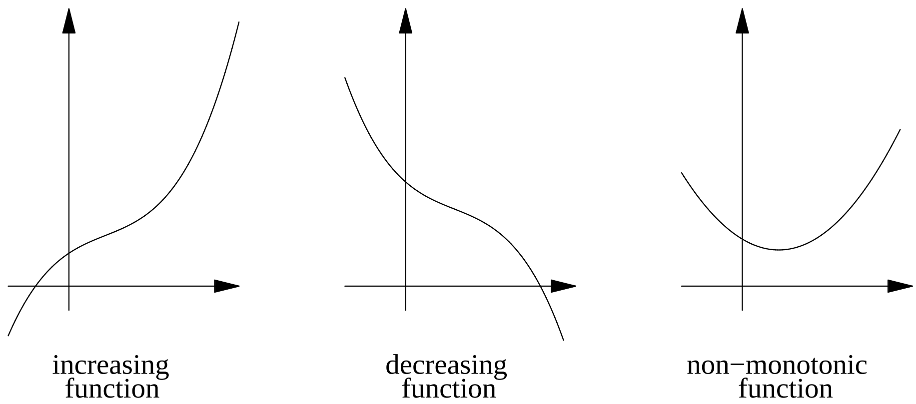
We say that f is *strictly increasing on J* if

$$\text{for all } x, y \text{ in } J\Big[(x < y) \implies (f(x) < f(y))\Big]$$

and we say that f is *strictly decreasing on J* if

$$\text{for all } x, y \text{ in } J\Big[(x < y) \implies (f(x) > f(y))\Big].$$

We say that f is *monotonic on J* if f is either increasing on J or decreasing on J, and we say that f is *strictly monotonic* on J if f is either strictly increasing or strictly decreasing on J.



A constant function on J is both increasing and decreasing on J.

5.33 Notation ($S_a^b f$) Let f be a function from the interval $[a, b]$ to the non-negative numbers. We will write

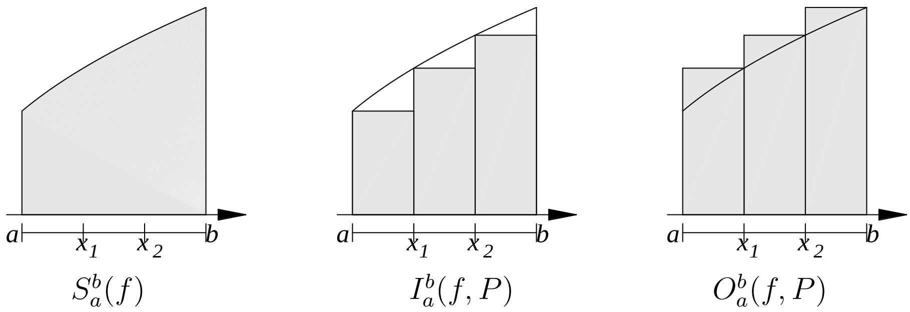
$$S_a^b f = \{(x, y) \in \mathbf{R}^2 \colon a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\},$$

i.e., $S_a^b f$ is the set of points under the graph of f over the interval $[a, b]$.

Let f be an increasing function from the interval $[a, b]$ to the non-negative numbers. Let $P = \{x_0, \cdots, x_n\}$ be a partition of $[a, b]$ and let

$$\begin{aligned} I_a^b(f, P) &= \bigcup_{i=1}^{n} B(x_{i-1}, x_i \colon 0, f(x_{i-1})) \\ O_a^b(f, P) &= \bigcup_{i=1}^{n} B(x_{i-1}, x_i \colon 0, f(x_i)). \end{aligned}$$

Then



$$I_a^b(f, P) \subset S_a^b f \subset O_a^b(f, P). \tag{5.34}$$

To see this, observe that since f is increasing

$$x_{i-1} \le x \le x_i \implies f(x_{i-1}) \le f(x) \le f(x_i),$$

so

$$\begin{aligned} (x, y) \in I_a^b(f, P) &\implies (x, y) \in B(x_{i-1}, x_i : 0, f(x_{i-1})) \text{ for some } i \\ &\implies x_{i-1} \le x \le x_i \text{ and } 0 \le y \le f(x_{i-1}) \le f(x) \text{ for some } i \\ &\implies a \le x \le b \text{ and } 0 \le y \le f(x) \implies (x, y) \in S_a^b f. \end{aligned}$$

and also

$$\begin{aligned} (x, y) \in S_a^b f &\implies x_{i-1} \le x \le x_i \text{ and } 0 \le y \le f(x) \le f(x_i) \text{ for some } i \\ &\implies (x, y) \in B\Big(x_{i-1}, x_i \colon 0, f(x_i)\Big) \text{ for some } i \\ &\implies (x, y) \in O_a^b(f, P). \end{aligned}$$

By equation (5.34) and monotonicity of area, we have

$$\alpha\Big(I_a^b(f,P)\Big) \le \alpha(S_a^b f) \le \alpha\Big(O_a^b(f,P)\Big). \tag{5.35}$$

Now

$$\begin{aligned}
&\alpha\Big(O_a^b(f,P)\Big) - \alpha\Big(I_a^b(f,P)\Big)\\
&= \sum_{i=1}^n (x_i - x_{i-1}) f(x_i) - \sum_{i=1}^n (x_i - x_{i-1}) f(x_{i-1})\\
&= \sum_{i=1}^n (x_i - x_{i-1})\Big(f(x_i) - f(x_{i-1})\Big). \qquad (5.36)
\end{aligned}$$

Now let $\mu(P)$ be the mesh of P (cf. definition 5.27) so that

$$0 \le x_i - x_{i-1} \le \mu(P) \text{ for } 1 \le i \le n.$$

Since $f(x_i) - f(x_{i-1}) \ge 0$ for all i, we have

$$(x_i - x_{i-1})\Big(f(x_i) - f(x_{i-1})\Big) \le \mu(P)\Big(f(x_i) - f(x_{i-1})\Big)$$

for all i, and hence

$$\begin{aligned}
\sum_{i=1}^n (x_i - x_{i-1})\Big(f(x_i) - f(x_{i-1})\Big) &\le \sum_{i=1}^n \mu(P)\Big(f(x_i) - f(x_{i-1})\Big)\\
&= \mu(P)\sum_{i=1}^n \Big(f(x_i) - f(x_{i-1})\Big). \qquad (5.37)
\end{aligned}$$

Now

$$\begin{aligned}
\sum_{i=1}^n \Big(f(x_i) - f(x_{i-1})\Big) &= \Big(f(x_n) - f(x_{n-1})\Big) + \Big(f(x_{n-1}) - f(x_{n-2})\Big)\\
&\quad + \cdots + \Big(f(x_1) - f(x_0)\Big)\\
&= f(x_n) - f(x_0) = f(b) - f(a).
\end{aligned}$$

so by equations (5.37) and (5.36), we have

$$\alpha\Big(O_a^b(f,P)\Big) - \alpha\Big(I_a^b(f,P)\Big) \le \mu(P)\Big(f(b) - f(a)\Big).$$

Now suppose that A is any real number that satisfies

$$\alpha(I_a^b(f,P)) \leq A \leq \alpha(O_a^b(f,P)) \text{ for every partition } P \text{ of } [a,b].$$

We will show that $A = \alpha(S_a^b f)$. We have

$$-\alpha(O_a^b(f,P)) \leq -A \leq -\alpha(I_a^b(f,P)).$$

It follows from (5.35) that

$$\alpha(I_a^b(f,P)) - \alpha(O_a^b(f,P)) \leq \alpha(S_a^b f) - A \leq \alpha(O_a^b(f,P)) - \alpha(I_a^b(f,P)).$$

Thus

$$-\mu(P)(f(b)-f(a)) \leq \alpha(S_a^b f) - A \leq \mu(P)(f(b)-f(a)) \tag{5.38}$$

for every partition P of $[a,b]$. Since we can find partitions P with $\mu(P)$ smaller than any preassigned number, it follows that

$$A = \alpha(S_a^b f). \tag{5.39}$$

(After we have discussed the notion of *limit*, we will come back and reconsider how (5.39) follows from (5.38). For the present, I will just say that the implication is intuitively clear.) We have now proved the following theorem:

5.40 Theorem. *Let f be an increasing function from the interval $[a,b]$ to $\mathbf{R}_{\geq 0}$, and let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a,b]$. Let*

$$S_a^b f = \{(x,y) \in \mathbf{R}^2 \colon a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\},$$

$$I_a^b(f,P) = \bigcup_{i=1}^{n} B\Big(x_{i-1}, x_i \colon 0, f(x_{i-1})\Big), \tag{5.41}$$

$$O_a^b(f,P) = \bigcup_{i=1}^{n} B\Big(x_{i-1}, x_i \colon 0, f(x_i)\Big), \tag{5.42}$$

$$A_a^b f = \alpha(S_a^b f).$$

Then

$$\alpha\Big(I_a^b(f,P)\Big) \leq A_a^b f \leq \alpha\Big(O_a^b(f,P)\Big) \tag{5.43}$$

and

$$\alpha\Big(O_a^b(f,P)\Big) - \alpha\Big(I_a^b(f,P)\Big) \leq \mu(P)\Big(f(b)-f(a)\Big). \tag{5.44}$$

Also

$$\alpha\Big(I_a^b(f,P)\Big) = \sum_{i=1}^{n} f(x_{i-1})(x_i - x_{i-1}) \qquad (5.45)$$

$$\alpha\Big(O_a^b(f,P)\Big) = \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1}). \qquad (5.46)$$

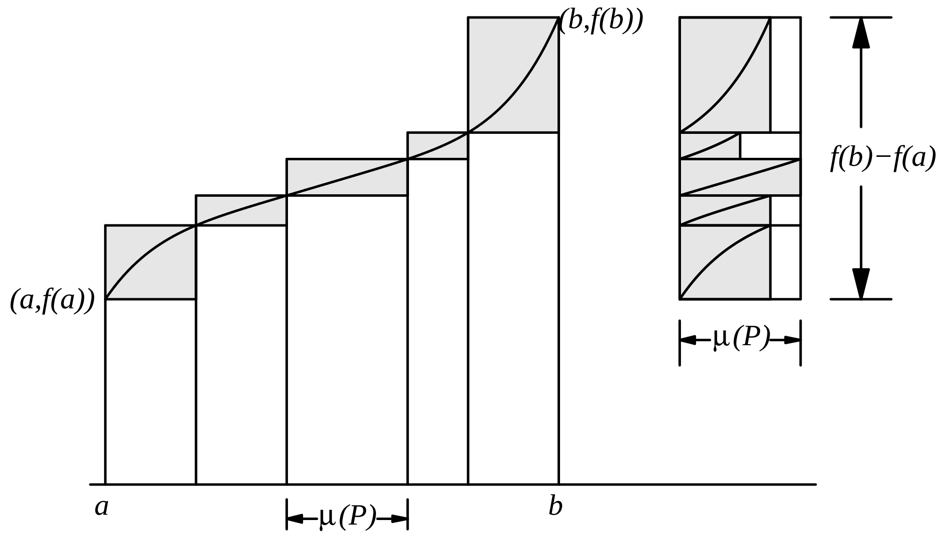
If A is any real number such that

$$\alpha(I_a^b(f,P)) \leq A \leq \alpha(O_a^b(f,P)) \text{ for every partition } P \text{ of } [a,b],$$

then

$$A = A_a^b(f).$$

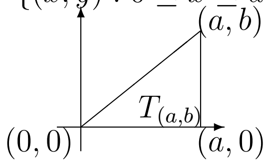
The following picture illustrates the previous theorem.



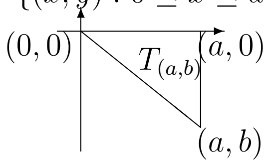
5.47 Exercise. A version of theorem 5.40 for decreasing functions is also valid. To get this version you should replace the word "increasing" by "decreasing" and change lines (5.41), (5.42), (5.44), (5.45) and (5.46). Write down the proper versions of the altered lines. As usual, use I to denote areas inside $S_a^b f$ and O to denote sets containing $S_a^b f$. Draw a picture corresponding to the above figure for a decreasing function.

5.48 Definition (Right triangle $T_{\mathbf{c}}$) Let a and b be non-zero real numbers, and let $\mathbf{c} = (a, b)$. We define the triangle $T_{\mathbf{c}} = T_{(a,b)}$ to be the set of points between the line segment $[\mathbf{0c}]$ and the x-axis. By example 4.8, we know that the equation of the line through $\mathbf{0}$ and $\mathbf{c}$ is $y = \frac{b}{a}x$. Hence we have:

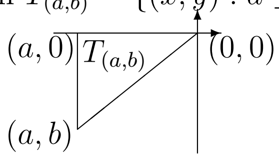
If $a > 0$ and $b > 0$, then $T_{(a,b)} = \{(x, y) : 0 \le x \le a \text{ and } 0 \le y \le \frac{b}{a}x\}$



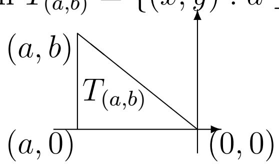
If $a > 0$ and $b < 0$, then $T_{(a,b)} = \{(x, y) : 0 \le x \le a \text{ and } \frac{b}{a}x \le y \le 0\}$



If $a < 0$ and $b < 0$, then $T_{(a,b)} = \{(x, y) : a \le x \le 0 \text{ and } \frac{b}{a}x \le y \le 0\}$



If $a < 0$ and $b > 0$, then $T_{(a,b)} = \{(x, y) : a \le x \le 0 \text{ and } 0 \le y \le \frac{b}{a}x\}$



5.49 Remark. We know from Euclidean geometry that

$$\alpha(T_{(a,b)}) = \frac{1}{2}|a||b|. \tag{5.50}$$

I would like to show that this relation follows from our assumptions about area. If H, V and R_π are the reflections and rotation defined in definition 4.9, then we can show without difficulty that for $a > 0$ and $b > 0$

$$\begin{aligned} T_{(-a,b)} &= H(T(a,b)), \\ T_{(a,-b)} &= V(T_{(a,b)}), \text{ and} \\ T_{(-a,-b)} &= R_\pi(T_{(a,b)}) \end{aligned}$$

so by invariance of area under symmetry,

$$\alpha(T_{(a,b)}) = \alpha(T_{(-a,b)}) = \alpha(T_{(a,-b)}) = \alpha(T_{(-a,-b)})$$

when a and b are positive. It follows that if we prove formula (5.50) when a and b are positive, then the formula holds in all cases. For example if a and b are positive, and we know that (5.50) holds when a and b are positive, we get

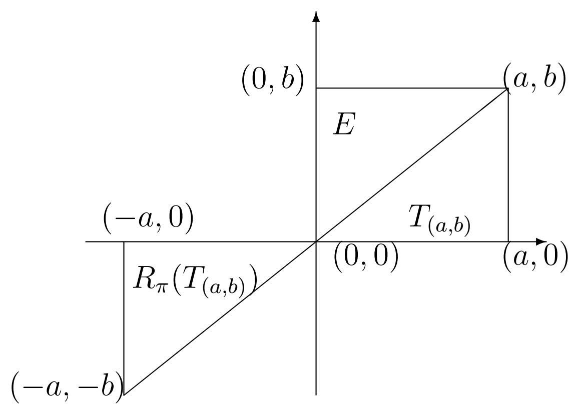
$$\alpha(T_{(-a,b)}) = \alpha(T_{(a,b)}) = \frac{1}{2}|a||b| = \frac{1}{2}|-a||b|,$$

and thus our formula holds when a is negative and b is positive.

5.51 Theorem. *Let a and b be non-zero real numbers, and let $T_{(a,b)}$ be the set defined in definition 5.48. Then*

$$\alpha(T_{(a,b)}) = \frac{1}{2}|a||b|.$$

Proof: By the previous remark, if is sufficient to prove the theorem for the case when a and b are positive. So suppose that a and b are positive.



Let $E = (a,b) + R_\pi(T_{(a,b)})$. It appears from the figure, and is straightforward to show, that

$$E = \{(x,y) : 0 \le x \le a \text{ and } \frac{b}{a}x \le y \le b\}.$$

By translation invariance of area,

$$\alpha(E) = \alpha(R_\pi(T_{(a,b)})) = \alpha(T_{(a,b)}).$$

We have

$$E \cup T_{(a,b)} = B(0, a : 0, b),$$

and

$$E \cap T_{(a,b)} = [\mathbf{0c}] \text{ where } \mathbf{c} = (a, b).$$

By the addition rule for area (assumption 5.14) we have

$$\begin{aligned} ab &= \alpha(B(0, a : 0, b)) \\ &= \alpha(E \cup T_{(a,b)}) \\ &= \alpha(E) + \alpha(T_{(a,b)}) - \alpha(E \cap T_{(a,b)}) \\ &= 2\alpha(T_{(a,b)}) - \alpha([\mathbf{0c}]), \end{aligned}$$

i.e.,

$$\alpha(T_{(a,b)}) = \frac{1}{2}ab + \frac{1}{2}\alpha([\mathbf{0c}]).$$

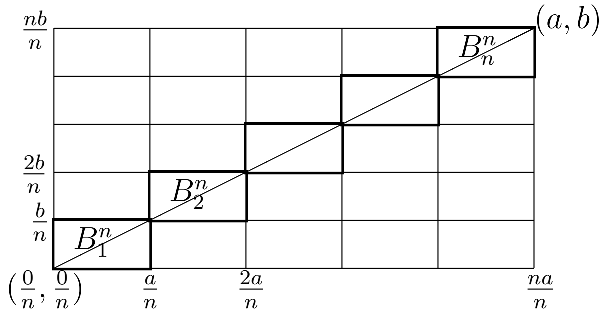
Thus our theorem will follow if we can show that the segment $[\mathbf{0c}]$ is a zero-area set. We will prove this as the next theorem.

5.52 Theorem. *Let* $\mathbf{c} = (a, b)$ *be a point in* $\mathbf{R}^2$. *Then*

$$\alpha([\mathbf{0c}]) = \mathbf{0}.$$

Proof: If $a = 0$ or $b = 0$, then $[\mathbf{0c}]$ is a box with width equal to zero, or height equal to zero, so the theorem holds in this case. Hence we only need to consider the case where a and b are non-zero. Since any segment $[\mathbf{0c}]$ can be rotated or reflected to a segment $[\mathbf{0q}]$ where $\mathbf{q}$ is in the first quadrant, we may further assume that a and b are both positive. Let n be a positive integer, and for $1 \leq j \leq n$ let

$$B_j^n = B\left(\frac{a(j-1)}{n}, \frac{aj}{n} : \frac{b(j-1)}{n}, \frac{bj}{n}\right).$$



Then

$$[\mathbf{0c}] \subset \bigcup_{j=1}^{n} B_j^n, \tag{5.53}$$

since

$$\begin{aligned}
\mathbf{x} \in [\mathbf{0c}] &\implies \mathbf{x} = (ta, tb) \text{ for some } t \in [0,1] \\
&\implies \mathbf{x} = (ta, tb) \text{ where } \frac{j-1}{n} \leq t \leq, \frac{j}{n} \text{ for some } j \text{ with } 1 \leq j \leq n \\
&\implies \mathbf{x} = (ta, tb) \text{ where } \frac{a(j-1)}{n} \leq ta \leq \frac{aj}{n} \\
&\qquad \text{and } \frac{b(j-1)}{n} \leq bt \leq \frac{bj}{n} \text{ for some } j \text{ with } 1 \leq j \leq n \\
&\implies \mathbf{x} \in B_j^n \text{ for some } j \text{ with } 1 \leq j \leq n.
\end{aligned}$$

For each j we have

$$\alpha(B_j^n) = \frac{a}{n} \cdot \frac{b}{n} = \frac{ab}{n^2}.$$

Also the sets B_j^n and B_k^n are almost disjoint whenever $1 \leq j, k \leq n$ and $j \neq k$. (If j and k differ by more than 1, then B_j^n and B_k^n are disjoint, and if j and k differ by 1, then $B_j^n \cap B_k^n$ consists of a single point.) By additivity for almost-disjoint sets (assumption 5.25), it follows that

$$\alpha(\bigcup_{j=1}^{n} B_j^n) = \sum_{j=1}^{n} \alpha(B_j^n) = \sum_{j=1}^{n} \frac{ab}{n^2} = \frac{nab}{n^2} = \frac{ab}{n}.$$

By (5.53) and monotonicity of area we have

$$\alpha([0\mathbf{c}]) \leq \alpha(\bigcup_{j=1}^{n} B_j^n) = \frac{ab}{n} \text{ for every positive integer } n. \qquad (5.54)$$

In order to conclude from this that $\alpha([\mathbf{0c}]) = 0$ We now make use of the *Archimedean property* of real numbers (see (C.79) in Appendix C) which says that for any real number x there is a positive integer n with $n > x$. We know $\alpha([\mathbf{0c}]) \geq 0$, since all areas are non-negative. Suppose (in order to get a contradiction) that $\alpha([\mathbf{0c}])$ is positive. Then by the Archimedean property, there is a positive integer N such that $N > \frac{ab}{\alpha([\mathbf{0c}])}$. This implies that $\alpha([\mathbf{0c}]) > \frac{ab}{N}$, and this contradicts (5.54). Hence $\alpha([\mathbf{0c}])$ is not positive, and we conclude that $\alpha([\mathbf{0c}]) = 0$. |||

Archimedes' statement of the Archimedean property differs from our statement. He assumes that

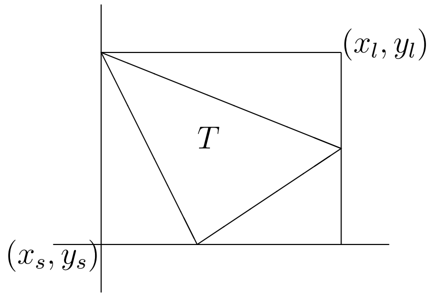
> Further, of unequal lines, unequal surfaces, and unequal solids, the greater exceeds the less by such a magnitude as, when added to itself, can be made to exceed any assigned magnitude among those which are comparable with [it and with] one another.[2, page 4]

5.55 Exercise. Let $\mathbf{a}$ and $\mathbf{b}$ be points in $\mathbf{R}^2$. Show that segment $[\mathbf{ab}]$ is a zero are set. (Use theorem 5.52. Do not reprove theorem 5.52).

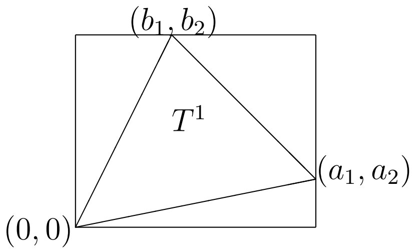
5.56 Entertainment (Area of a triangle) Let $\mathbf{x}_1 = (x_1, y_1)$, $\mathbf{x}_2 = (x_2, y_2)$ and $\mathbf{x}_3 = (x_3, y_3)$ be three points in $\mathbf{R}^2$, and let T be the triangle with vertices $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. Let

$$\begin{aligned} x_s &= \text{smallest of } x_1, x_2 \text{ and } x_3 \\ x_l &= \text{largest of } x_1, x_2 \text{ and } x_3 \\ y_s &= \text{smallest of } y_1, y_2 \text{ and } y_3 \\ y_l &= \text{largest of } y_1, y_2 \text{ and } y_3. \end{aligned}$$

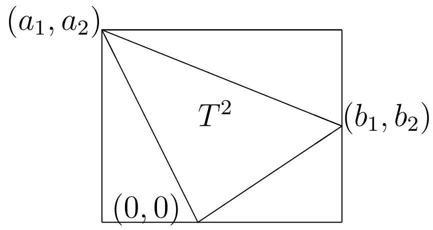
Then the box $B(x_s, x_l : y_s, y_l)$ is an almost-disjoint union of T and three triangles which are translates of triangles of the form $T_{\mathbf{c}}$. Since you know how to find the area of a box and of a triangle $T_{\mathbf{c}}$, you can find the area of T.



Using this remark show that for the triangles pictured below, $\alpha(T^1) = \frac{1}{2}(a_1b_2 - a_2b_1)$, and $\alpha(T^2) = \frac{1}{2}(a_2b_1 - a_1b_2)$.



$0 < b_1 < a_1$ and $0 < b_2 < a_2$



$a_1 < 0 < b_1$ and $0 < b_2 < a_2$

Then choose another triangle T^3 with vertices $\mathbf{0}, \mathbf{a}$ and $\mathbf{b}$, where the coordinates of the points are related in a way different from the ways shown for T^1 and T^2, and calculate the area of T^3. You should find that

$$\alpha(T^3) = \frac{1}{2}|a_1b_2 - a_2b_1|$$

in all cases. Notice that if some coordinate is zero, the formula agrees with theorem 5.51.

5.4 Logarithms.

5.57 Notation ($A_a^b f$, $A_a^b[f(t)]$.) Let f be a bounded function from the interval $[a, b]$ to $\mathbf{R}_{\geq 0}$. We will denote the area of $S_a^b f$ by $A_a^b f$. Thus

$$A_a^b f = \alpha\Big(\{(x, y) \in \mathbf{R}^2 \colon a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}\Big)$$

We will sometimes write $A_a^b[f(t)]$ instead of $A_a^b f$. Thus, for example

$$A_a^b[t^2] = \alpha\Big(\{(x,y) \in \mathbf{R}^2 \colon a \le x \le b \text{ and } 0 \le y \le x^2\}\Big)$$

We will also write $I_a^b([f(t)], P)$ and $O_a^b([f(t)], P)$ for $I_a^b(f, P)$ and $O_a^b(f, P)$ respectively.

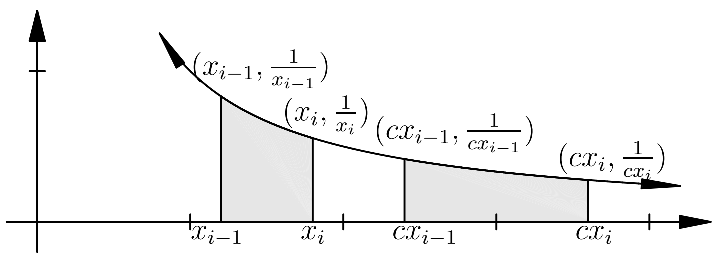
5.58 Lemma. [2] *Let a, b, and c be real numbers such that $0 < a < b$ and $c > 0$. Then*

$$A_{ac}^{bc}\Big[\frac{1}{t}\Big] = A_a^b\Big[\frac{1}{t}\Big].$$

Proof: Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a, b]$, and let

$$cP = \{cx_0, cx_1, \cdots, cx_n\}$$

be the partition of $[ca, cb]$ obtained by multiplying the points of P by c.



Then

$$\begin{aligned} \alpha(I_{ac}^{bc}(\Big[\frac{1}{t}\Big], cP)) &= \sum_{i=1}^{n} \frac{1}{cx_i}(cx_i - cx_{i-1}) = \sum_{i=1}^{n} \frac{1}{cx_i} \cdot c(x_i - x_{i-1}) \\ &= \sum_{i=1}^{n} \frac{1}{x_i}(x_i - x_{i-1}) = \alpha(I_a^b(\Big[\frac{1}{t}\Big], P)) \qquad (5.59) \end{aligned}$$

and

$$\begin{aligned} \alpha(O_{ac}^{bc}(\Big[\frac{1}{t}\Big], cP)) &= \sum_{i=1}^{n} \frac{1}{cx_{i-1}}(cx_i - cx_{i-1}) = \sum_{i=1}^{n} \frac{1}{cx_{i-1}} \cdot c(x_i - x_{i-1}) \\ &= \sum_{i=1}^{n} \frac{1}{x_{i-1}}(x_i - x_{i-1}) = \alpha(O_a^b(\Big[\frac{1}{t}\Big], P)) \qquad (5.60) \end{aligned}$$

[2] A lemma is a theorem which is proved in order to help prove some other theorem.

We know that

$$\alpha(I_{ac}^{bc}(\left[\frac{1}{t}\right], cP)) \leq A_{ac}^{bc}\left[\frac{1}{t}\right] \leq \alpha(O_{ac}^{bc}(\left[\frac{1}{t}\right], cP)).$$

Hence by (5.59) and (5.60) we have

$$\alpha(I_{a}^{b}(\left[\frac{1}{t}\right], P)) \leq A_{ac}^{bc}\left[\frac{1}{t}\right] \leq \alpha(O_{a}^{b}(\left[\frac{1}{t}\right], P))$$

for every partition P of $[a, b]$. It follows from this and the last statement of theorem 5.40 that

$$A_{ac}^{bc}\left[\frac{1}{t}\right] = A_{a}^{b}\left[\frac{1}{t}\right]. \;\|$$

5.61 Exercise. From lemma 5.58 we see that

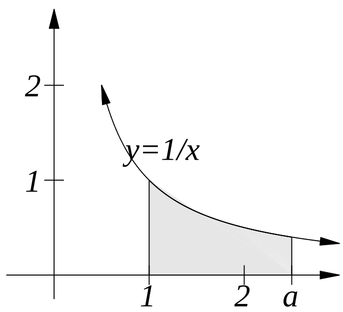
$$A_{a}^{b}\left[\frac{1}{t}\right] = A_{ac}^{bc}\left[\frac{1}{t}\right]$$

whenever $0 < a < b$, and $c > 0$. Use this result to show that for $a \geq 1$ and $b \geq 1$

$$A_{1}^{ab}\left[\frac{1}{t}\right] = A_{1}^{a}\left[\frac{1}{t}\right] + A_{1}^{b}\left[\frac{1}{t}\right]. \tag{5.62}$$

5.63 Definition ($L(x)$.) We will define a function $L\colon [1, \infty) \to \mathbf{R}$ by

$$L(a) = A_{1}^{a}\left[\frac{1}{t}\right] \text{ for all } a \in [1, \infty).$$



shaded area = $L(a)$

By exercise 5.61 we have

$$L(ab) = L(a) + L(b) \text{ for all } a \geq 1,\ b \geq 1. \tag{5.64}$$

In this section we will extend the domain of L to all of $\mathbf{R}^+$ in such a way that (5.64) holds for all $a, b \in \mathbf{R}^+$.

5.65 Theorem. *Let a, b, c be real numbers such that $a \leq b \leq c$, and let f be a bounded function from $[a, b]$ to $\mathbf{R}_{\geq 0}$. Then*

$$A_a^c f = A_a^b f + A_b^c f. \tag{5.66}$$

Proof: We want to show

$$\alpha(S_a^c f) = \alpha(S_a^b f) + \alpha(S_b^c f).$$

Since $S_a^c f = S_a^b f \cup S_b^c f$ and the sets $S_a^b f$ and $S_b^c f$ are almost disjoint, this conclusion follows from our assumption about additivity of area for almost disjoint sets.

I now want to extend the definition of $A_a^b f$ to cases where b may be less than a. I want equation (5.66) to continue to hold in all cases. If $c = a$ in (5.66), we get

$$0 = A_a^a f = A_a^b f + A_b^a f$$

i.e.,

$$A_b^a f = -A_a^b f.$$

Thus we make the following definition:

5.67 Definition. Let a, b be real numbers with $a \leq b$ and let f be a bounded function from $[a, b]$ to $\mathbf{R}_{\geq 0}$. Then we define

$$A_b^a f = -A_a^b f \text{ or } A_b^a[f(t)] = -A_a^b[f(t)].$$

5.68 Theorem. *Let a, b, c be real numbers and let f be a bounded non-negative real valued function whose domain contains an interval containing a, b, and c. Then*

$$A_a^c f = A_a^b f + A_b^c f.$$

Proof: We need to consider the six possible orderings for a, b and c. If $a \le b \le c$ we already know the result. Suppose $b \le c \le a$. Then $A_b^a f = A_b^c f + A_c^a f$ and hence $-A_a^b f = A_b^c f - A_a^c f$, i.e., $A_a^c f = A_a^b f + A_b^c f$. The remaining four cases are left as an exercise.

5.69 Exercise. Prove the remaining four cases of theorem 5.68.

5.70 Definition (Logarithm.) If a is any positive number, we define the *logarithm of* a by

$$\ln(a) = L(a) = A_1^a\Big[\frac{1}{t}\Big].$$

5.71 Theorem (Properties of Logarithms.) *For all $a, b \in \mathbf{R}^+$ and all $r \in \mathbf{Q}$ we have*

$$\begin{aligned}
L(ab) &= L(a) + L(b) \\
L\Big(\frac{a}{b}\Big) &= L(a) - L(b) \\
L(a^{-1}) &= -L(a) \\
L(a^r) &= rL(a) \qquad (5.72)\\
L(1) &= 0. \qquad (5.73)
\end{aligned}$$

Proof: Let $a, b, c \in \mathbf{R}^+$. From lemma 5.58 we know that if $a \le c$ then

$$A_a^c\Big[\frac{1}{t}\Big] = A_{ba}^{bc}\Big[\frac{1}{t}\Big] \qquad (5.74)$$

If $c < a$ we get

$$A_a^c\Big[\frac{1}{t}\Big] = -A_c^a\Big[\frac{1}{t}\Big] = -A_{bc}^{ba}\Big[\frac{1}{t}\Big] = A_{ba}^{bc}\Big[\frac{1}{t}\Big]$$

so equation (5.74) holds in all cases. Let a, b be arbitrary elements in $\mathbf{R}^+$. Then

$$\begin{aligned}
L(ab) &= A_1^{ab}\Big[\frac{1}{t}\Big] = A_1^a\Big[\frac{1}{t}\Big] + A_a^{ab}\Big[\frac{1}{t}\Big] \\
&= A_1^a\Big[\frac{1}{t}\Big] + A_1^b\Big[\frac{1}{t}\Big] = L(a) + L(b).
\end{aligned}$$

Also

$$L(1) = A_1^1\Big[\frac{1}{t}\Big] = 0,$$

so

$$0 = L(1) = L(a \cdot a^{-1}) = L(a) + L(a^{-1})$$

and it follows from this that

$$L(a^{-1}) = -L(a).$$

Hence

$$L\Big(\frac{a}{b}\Big) = L(a \cdot b^{-1}) = L(a) + L(b^{-1}) = L(a) - L(b).$$

5.75 Lemma. *For all $n \in \mathbf{Z}_{\geq 0}$, $L(a^n) = nL(a)$.*

Proof: The proof is by induction on n. For $n = 0$ the lemma is clear. Suppose now that the lemma holds for some $n \in \mathbf{Z}_{\geq 0}$, i.e., suppose that $L(a^n) = nL(a)$. Then

$$L(a^{n+1}) = L(a^n \cdot a) = L(a^n) + L(a) = nL(a) + L(a) = (n+1)L(a).$$

The lemma now follows by induction.

If $n \in \mathbf{Z}^-$ then $-n \in \mathbf{Z}^+$ and

$$L(a^n) = L\Big((a^{-n})^{-1}\Big) = -L(a^{-n}) = -(-n)L(a) = nL(a).$$

Thus equation (5.72) holds whenever $r \in \mathbf{Z}$. If $p \in \mathbf{Z}$ and $n \in \mathbf{Z} \setminus \{0\}$, then

$$pL(a) = L(a^p) = L\Big((a^{\frac{p}{n}})^n\Big) = nL\Big(a^{\frac{p}{n}}\Big)$$

so

$$L\Big(a^{\frac{p}{n}}\Big) = \frac{p}{n}L(a).$$

Thus (5.72) holds for all $r \in Q$. |||

5.76 Theorem. *Let a and b be numbers such that $0 < a < b$. Then*

$$A_a^b\Big[\frac{1}{t}\Big] = \ln\Big(\frac{b}{a}\Big) = \ln(b) - \ln(a).$$

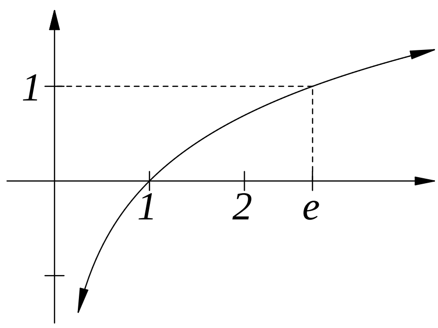
Proof: By lemma 5.58

$$A_a^b\left[\frac{1}{t}\right] = A_{aa^{-1}}^{ba^{-1}}\left[\frac{1}{t}\right] = A_1^{ba^{-1}}\left[\frac{1}{t}\right] = \ln(\frac{b}{a}) = \ln(b) - \ln(a). \;\|$$

Logarithms were first introduced by John Napier (1550-1632) in 1614. Napier made up the word *logarithm* from Greek roots meaning *ratio number*, and he spent about twenty years making tables of them. As far as I have been able to find out, the earliest use of ln for logarithms was by Irving Stringham in 1893[15, vol 2, page 107]. The notation $\log(x)$ is probably more common among mathematicians than $\ln(x)$, but since calculators almost always calculate our function with a key called "ln", and calculate something else with a key called "log", I have adopted the "ln" notation. (Napier did not use any abbreviation for logarithm.) Logarithms were seen as an important computational device for reducing multiplications to additions. The first explicit notice of the fact that logarithms are the same as areas of hyperbolic segments was made in 1649 by Alfons Anton de Sarasa (1618-1667), and this observation increased interest in the problem of calculating areas of hyperbolic segments.

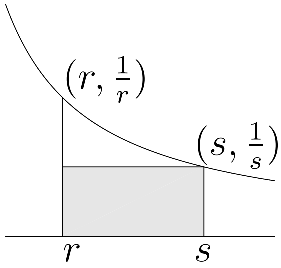
5.77 Entertainment (Calculate $\ln(2)$.**)** Using any computer or calculator, compute $\ln(2)$ accurate to 10 decimal places. You should not make use of any special functions, e.g., it is not fair to use the "ln" key on your calculator. There are better polygonal approximations to $A_1^2\left[\frac{1}{t}\right]$ than the ones we have discussed.

The graph of the logarithm function is shown below.



y= ln(*x*)

We know that $\ln(1) = 0$ and it is clear that ln is strictly increasing.



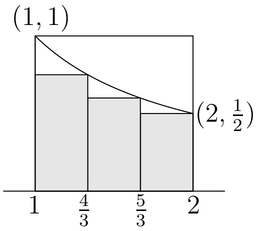
If $0 < r < s$, then

$$\ln(s) - \ln(r) = A_r^s\Big[\frac{1}{t}\Big] > (s-r)\frac{1}{s} > 0.$$

From the fact that $\ln(a^n) = n\ln(a)$ for all $n \in \mathbf{Z}$, it is clear that ln takes on arbitrarily large positive and negative values, but the function increases very slowly. Let

$$P = \{1, \frac{4}{3}, \frac{5}{3}, \frac{6}{3}\}$$

be the regular partition of $[1, 2]$ into three subintervals.



Then

$$\begin{aligned} \ln(2) &= A_1^2\Big[\frac{1}{t}\Big] \geq \alpha(I_1^2(\Big[\frac{1}{t}\Big], P)) \\ &= \frac{1}{3}\cdot\frac{3}{4} + \frac{1}{3}\cdot\frac{3}{5} + \frac{1}{3}\cdot\frac{3}{6} = \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = \frac{37}{60}. \end{aligned}$$

Now

$$\ln(2) = A_1^2\Big[\frac{1}{t}\Big] \leq \alpha\Big(B(1, 2\colon 0, 1)\Big) = 1,$$

and

$$\ln(4) = \ln(2^2) = 2\ln(2) \geq 2 \cdot \frac{37}{60} > 1,$$

i.e.,

$$\ln(2) \leq 1 \leq \ln(4). \tag{5.78}$$

There is a unique number $e \in [2,4]$ such that $\ln(e) = 1$. The uniqueness is clear because ln is strictly increasing.

The existence of such a number was taken as obvious before the nineteenth century. Later we will introduce the *intermediate value property* which will allow us to prove that such a number e exists. For the time being, we will behave like eighteenth century mathematicians, and just assert that such a number e exists.

5.79 Definition (e.) We denote the unique number in $\mathbf{R}^+$ whose logarithm is 1 by e.

5.80 Exercise. Prove that $2 \leq e \leq 3$. (We already know $2 \leq e$.)

5.81 Entertainment (Calculate e.) Using any computing power you have, calculate e as accurately as you can, e.g., as a start, find the first digit after the decimal point.

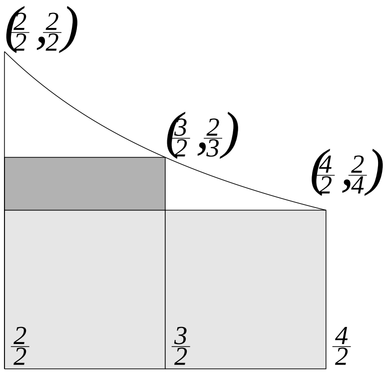
5.5 *Brouncker's Formula For $\ln(2)$

The following calculation of $\ln(2)$ is due to William Brouncker (1620-1684)[22, page 54].

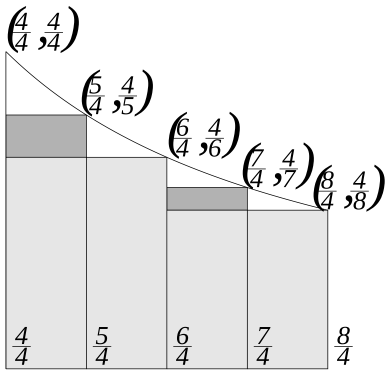
Let $P_{2^n} = \{x_0, x_1, \cdots, x_{2^n}\}$ denote the regular partition of the interval $[1,2]$ into 2^n equal subintervals. Let

$$K(2^n) = I_1^2(\Big[\frac{1}{t}\Big], P_{2^n}) = \bigcup_{i=1}^{2^n} B(x_{i-1}, x_i; 0, \frac{1}{x_i}).$$

We can construct $K(2^{n+1})$ from $K(2^n)$ by adjoining a box of width $\dfrac{1}{2^{n+1}}$ to the top of each box $B(x_{i-1}, x_i; 0, \frac{1}{x_i})$ that occurs in the definition of $K(2^n)$ (see figures a) and b)).



$K(1)$ = lightly shaded region
$K(2)$ = total shaded region
figure a



$K(2)$ = lightly shaded region
$K(4)$ = total shaded region
figure b

We have

$$\alpha(K(1)) = \alpha(B(1,2;0,\frac{1}{2})) = 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

From figure a) we see that

$$\begin{aligned}
\alpha(K(2)) &= \alpha(K(1)) + \alpha(B(\frac{2}{2},\frac{3}{2};\frac{2}{4},\frac{2}{3})) \\
&= \frac{1}{2} + \frac{1}{2}\Big(\frac{2}{3} - \frac{2}{4}\Big) \\
&= \frac{1}{2} + \Big(\frac{1}{3} - \frac{1}{4}\Big) \\
&= \frac{1}{2} + \frac{1}{3 \cdot 4}.
\end{aligned}$$

From figure b) we see that

$$\begin{aligned}
\alpha(K(4)) &= \alpha(K(2)) + \alpha(B(\frac{4}{4},\frac{5}{4};\frac{4}{6},\frac{4}{5})) + \alpha(B(\frac{6}{4},\frac{7}{4};\frac{4}{8},\frac{4}{7})) \\
&= \alpha(K(2)) + \frac{1}{4}\Big(\frac{4}{5} - \frac{4}{6}\Big) + \frac{1}{4}\Big(\frac{4}{7} - \frac{4}{8}\Big) \\
&= \alpha(K(2)) + \Big(\frac{1}{5} - \frac{1}{6}\Big) + \Big(\frac{1}{7} - \frac{1}{8}\Big) \\
&= \frac{1}{1 \cdot 2} + \frac{1}{3 \cdot 4} + \frac{1}{5 \cdot 6} + \frac{1}{7 \cdot 8}.
\end{aligned}$$

In general we will find that

$$\alpha(K(2^n)) = \sum_{j=1}^{2^n} \frac{1}{(2j-1)(2j)}.$$

Now

$$0 \le \alpha(S_1^2([\frac{1}{t}])) - \alpha(K(2^n)) \le (1 - \frac{1}{2})\mu(P_{2^n}),$$

i.e.

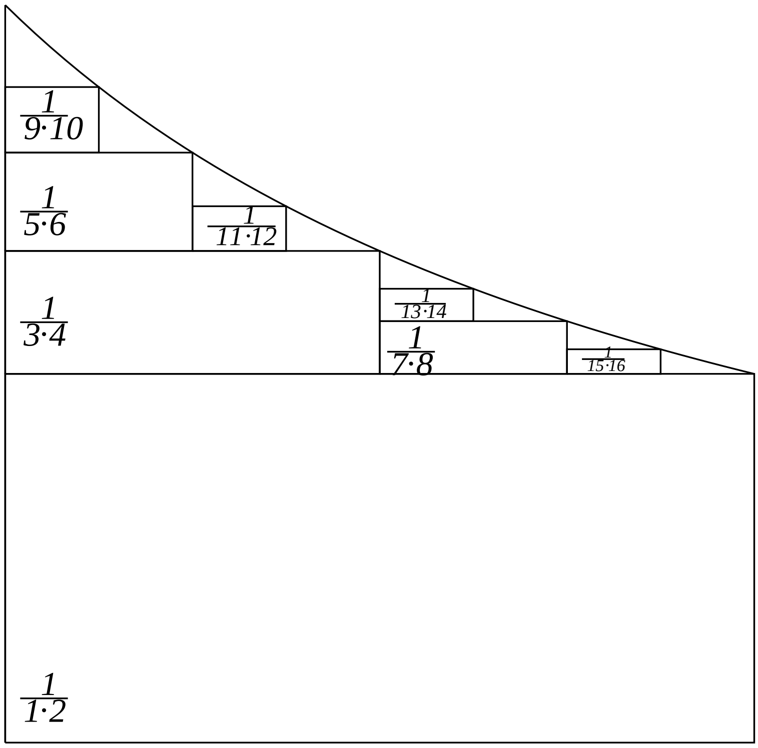
$$0 \le \ln(2) - \sum_{j=1}^{2^n} \frac{1}{(2j-1)(2j)} \le \frac{1}{2^{n+1}}.$$

Thus

$$\ln(2) = \sum_{j=1}^{2^n} \frac{1}{(2j-1)(2j)} \text{ with an error smaller than } \frac{1}{2^{n+1}}.$$

We can think of ln(2) as being given by the "infinite sum"

$$\ln(2) = \frac{1}{1\cdot 2} + \frac{1}{3\cdot 4} + \frac{1}{5\cdot 6} + \frac{1}{7\cdot 8} + \cdots. \tag{5.82}$$



$$\ln(2) = \tfrac{1}{1\cdot 2} + \tfrac{1}{3\cdot 4} + \tfrac{1}{5\cdot 6} + \cdots +$$

Equation (5.82) is sometimes called Mercator's expansion for ln(2), after Nicolaus Mercator, who found the result sometime near 1667 by an entirely different method.

Brouncker's calculation was published in 1668, but was done about ten years earlier [22, pages 56-56].

Brouncker's formula above is an elegant result, but it is not very useful for calculating: it takes too many terms in the sum to get much accuracy. Today, when a logarithm can be found by pressing a button on a calculator, we tend to think of "ln(2)" as being a known number, and of Brouncker's formula as giving a "closed form" for the sum of the infinite series $\dfrac{1}{1\cdot 2} + \dfrac{1}{3\cdot 4} + \dfrac{1}{5\cdot 6} + \cdots$.

5.6 Computer Calculation of Area

In this section we will discuss a Maple program for calculating approximate values of $A_a^b f$ for monotonic functions f on the interval $[a,b]$. The programs will be based on formulas discussed in theorem 5.40.

Let f be a decreasing function from the interval $[a,b]$ to $\mathbf{R}_{\geq 0}$, and let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition of $[a,b]$. We know that

$$\alpha(I_a^b(f,P)) \leq A_a^b f \leq \alpha(O_a^b(f,P)),$$

where

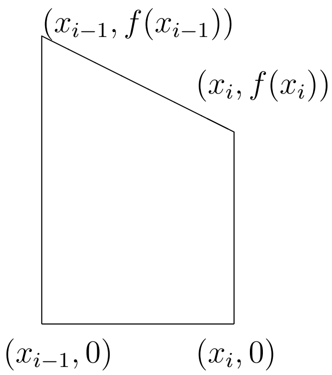
$$\alpha(I_a^b(f,P)) = \sum_{i=1}^{n}(x_i - x_{i-1})f(x_i), \tag{5.83}$$

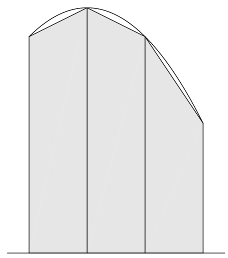
$$\alpha(O_a^b(f,P)) = \sum_{i=1}^{n}(x_i - x_{i-1})f(x_{i-1}). \tag{5.84}$$

Let $V_a^b(f,P)$ be the average of $\alpha(I_a^b(f,P))$ and $\alpha(O_a^b(f,P))$, so

$$V_a^b(f,P) = \frac{\alpha(I_a^b(f,P)) + \alpha(O_a^b(f,P))}{2} = \sum_{i=1}^{n}(x_i - x_{i-1}) \cdot \frac{f(x_i) + f(x_{i-1})}{2}.$$

Now $(x_i - x_{i-1}) \cdot \dfrac{f(x_i) + f(x_{x-1})}{2}$ represents the area of the trapezoid with vertices $(x_{i-1}, 0)$, $(x_{i-1}, f(x_{i-1}))$, $(x_i, f(x_i))$ and $(x_i, 0)$, so $V_a^b(f,P)$ represents the area under the polygonal line obtained by joining the points $(x_{i-1}, f(x_{i-1}))$ and $(x_i, f(x_i))$ for $1 \leq i \leq n$.





$V_a^b(f, P)$

In the programs below, `leftsum(f,a,b,n)` calculates

$$\sum_{j=1}^{n} f\left(a + (j-1)\Big(\frac{b-a}{n}\Big)\right)\Big(\frac{b-a}{n}\Big) = \Big(\frac{b-a}{n}\Big)\sum_{j=1}^{n} f\left(a + (j-1)\Big(\frac{b-a}{n}\Big)\right),$$

which corresponds to (5.84) when P is the regular partition of $[a, b]$ into n equal subintervals, and `rightsum(f,a,b,n)` calculates

$$\sum_{j=1}^{n} f\left(a + j\Big(\frac{b-a}{n}\Big)\right)\Big(\frac{b-a}{n}\Big) = \Big(\frac{b-a}{n}\Big)\sum_{j=1}^{n} f\left(a + j\Big(\frac{b-a}{n}\Big)\right).$$

which similarly corresponds to (5.83). The command `average(f,a,b,n)` calculates the average of `leftsum(f,a,b,n)` and `rightsum(f,a,b,n)`.

The equation of the unit circle is $x^2 + y^2 = 1$, so the upper unit semicircle is the graph of f where $f(x) = \sqrt{1 - x^2}$. The area of the unit circle is 4 times the area of the portion of the circle in the first quadrant, so

$$\pi = 4A_0^1[\sqrt{1 - t^2}].$$

Also

$$\ln(2) = A_1^2\Big[\frac{1}{t}\Big].$$

My routines and calculations are given below. Here `leftsum`, `rightsum` and `average` are all procedures with four arguments, `f,a,b,` and `n`.

`f` is a function.

`a` and `b` are the endpoints of an interval.

`n` is the number of subintervals in a partition of `[a,b]`.

The functions `F` and `G` are defined by `F`$(x) = 1/x$ and `G`$(x) = \sqrt{1-x^2}$. The command

```
average(F,1.,2.,10000);
```

estimates $\ln(2)$ by considering the regular partition of $[1,2]$ into 10000 equal subintervals. and the command

```
4*average(G,0.,1.,2000);
```

estimates π by considering the regular partition of $[0,1]$ into 2000 equal subintervals.

```
> leftsum :=
>       (f,a,b,n) -> (b-a)/n*sum(f( a +((j-1)*(b-a))/n),j=1..n);
```

$$leftsum := (f, a, b, n) \to \frac{(b-a)\left(\sum_{j=1}^{n} f\left(a + \frac{(j-1)(b-a)}{n}\right)\right)}{n}$$

```
> rightsum :=
>       (f,a,b,n) -> (b-a)/n*sum(f( a +(j*(b-a))/n),j=1..n);
```

$$rightsum := (f, a, b, n) \to \frac{(b-a)\left(\sum_{j=1}^{n} f\left(a + \frac{j(b-a)}{n}\right)\right)}{n}$$

```
> average :=
>       (f,a,b,n) -> (leftsum(f,a,b,n) + rightsum(f,a,b,n))/2;
```

$$average := (f, a, b, n) \to \frac{1}{2}\,\mathrm{leftsum}(f, a, b, n) + \frac{1}{2}\,\mathrm{rightsum}(f, a, b, n)$$

```
> F := t -> 1/t;
```

$$F := t \to \frac{1}{t}$$

```
> leftsum(F,1.,2.,10000);
```

$$.6931721810$$

```
> rightsum(F,1.,2.,10000);
```

$$.6931221810$$

```
> average(F,1.,2.,10000);
```

$$.6931471810$$

```
> ln(2.);
```

$$.6931471806$$

```
> G := t -> sqrt(1-t^2);
```

$$G := t \to \mathrm{sqrt}(\,1 - t^2\,)$$

```
> 4*leftsum(G,0.,1.,2000);
```

$$3.142579520$$

```
> 4*rightsum(G,0.,1.,2000);
```

$$3.140579522$$

```
> 4*average(G,0.,1.,2000);
```

$$3.141579521$$

```
> evalf(Pi);
```

$$3.141592654$$

Observe that in these examples, `average` yields much more accurate approximations than either `leftsum` or `rightsum`.

Chapter 6

Limits of Sequences

6.1 Absolute Value

6.1 Definition (Absolute values.) Recall that if x is a real number, then the *absolute value of* x, denoted by $|x|$, is defined by

$$|x| = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x < 0. \end{cases}$$

We will assume the following properties of absolute value, that follow easily from the definition:

For all real numbers x, y, z with $z \neq 0$

$$\begin{aligned} |x| &= |-x| \\ |xy| &= |x| \cdot |y| \\ \left|\frac{x}{z}\right| &= \frac{|x|}{|z|} \\ -|x| &\leq x \leq |x|. \end{aligned}$$

For all real numbers x, and all $a \in \mathbf{R}^+$

$$(|x| < a) \iff (-a < x < a)$$

and

$$(|x| \leq a) \iff (-a \leq x \leq a). \tag{6.2}$$

We also have

$$|x| \in \mathbf{R}_{\geq 0} \text{ for all } x \in \mathbf{R},$$

and

$$|x| = 0 \iff x = 0.$$

6.3 Theorem. *Let $a \in \mathbf{R}$ and let $p \in \mathbf{R}^+$. Then for all $x \in \mathbf{R}$ we have*

$$|x - a| < p \iff (a - p < x < a + p),$$

and

$$|x - a| \leq p \iff (a - p \leq x \leq a + p).$$

Equivalently, we can say that

$$\{x \in \mathbf{R} : |x - a| < p\} = (a - p, a + p)$$

and

$$\{x \in \mathbf{R} : |x - a| \leq p\} = [a - p, a + p].$$

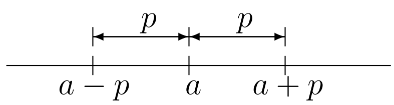
Proof: I will prove only the first statement. I have

$$\begin{aligned} |x - a| < p &\iff -p < x - a < p \\ &\iff a - p < a + (x - a) < a + p \\ &\iff a - p < x < a + p. \ \| \end{aligned}$$

6.4 Definition (Distance.) The *distance* between two real numbers x and y is defined by

$$\text{dist}(x, y) = |x - y|.$$

Theorem 6.3 says that the set of numbers whose distance from a is smaller than p is the interval $(a-p, a+p)$. Geometrically this is clear from the picture.



I remember the theorem by keeping the picture in mind.

6.5 Theorem (Triangle inequality.) *For all real numbers x and y*

$$|x + y| \leq |x| + |y|, \tag{6.6}$$

Proof For all x and y in $\mathbf{R}$ we have

$$-|x| \leq x \leq |x|$$

and

$$-|y| \leq y \leq |y|,$$

so

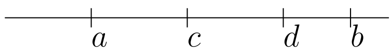
$$-(|x| + |y|) \leq x + y \leq (|x| + |y|).$$

Hence (Cf. (6.2))

$$|x + y| \leq |x| + |y|.$$

6.7 Exercise. Can you prove that for all $(x, y) \in \mathbf{R}^2 \Big(|x - y| \leq |x| - |y|\Big)$? Can you prove that for all $(x, y) \in \mathbf{R}^2 \Big(|x - y| \leq |x| + |y|\Big)$?

Remark: Let a, b, c, d be real numbers with $a < c < b$ and $a < d < b$.



Then

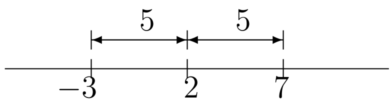
$$|c - d| < |b - a| = b - a.$$

This result should be clear from the picture. We can give an analytic proof as follows.

$$\begin{aligned}
(a < c < b \text{ and } a < d < b) &\implies (a < c < b \text{ and } -b < -d < -a) \\
&\implies a - b < c - d < b - a \\
&\implies -(b - a) < c - d < (b - a) \\
&\implies -|b - a| < c - d < |b - a| \\
&\implies |c - d| < |b - a|.
\end{aligned}$$

6.8 Examples. Let

$$\begin{aligned}
A &= \{x \in \mathbf{R} : |x - 2| < 5\}, \\
B &= \{x \in \mathbf{R} : |x - 2| > 5\}.
\end{aligned}$$



Then a number x is in A if and only if the distance from x to 2 is smaller than 5, and x is in B if and only if the distance from x to 2 is greater than 5. I can see by inspection that

$$A = (-3, 7),$$

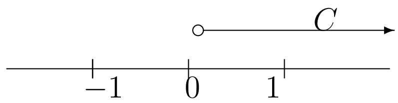
and

$$B = (-\infty, -3) \cup (7, \infty).$$

Let

$$C = \{x \in \mathbf{R} : \left|\frac{x-1}{x+1}\right| < 1\}.$$

If $x \in \mathbf{R} \setminus \{-1\}$, then x is in C if and only if $|x-1| < |x+1|$, i.e. if and only if x is closer to 1 than to -1.



I can see by inspection that the point equidistant from -1 and 1 is 0, and that the numbers that are closer to 1 than to -1 are the positive numbers, so $C = (0, \infty)$. I can also do this analytically, (but in practice I wouldn't) as follows. Since the alsolute values are all non-negative

$$\begin{aligned} |x-1| < |x+1| &\iff |x-1|^2 < |x+1|^2 \\ &\iff x^2 - 2x + 1 < x^2 + 2x + 1 \\ &\iff 0 < 4x \iff 0 < x. \end{aligned}$$

6.9 Exercise. Express each of the four sets below as an interval or a union of intervals. (You can do this problem by inspection.)

$$\begin{aligned} A_1 &= \{x \in \mathbf{R}\colon |x - \frac{1}{2}| < \frac{3}{2}\}, \\ A_2 &= \{x \in \mathbf{R}\colon |x + \frac{1}{2}| \le \frac{3}{2}\}, \\ A_3 &= \{x \in \mathbf{R}\colon |\frac{3}{2} - x| < \frac{1}{2}\}, \\ A_4 &= \{x \in \mathbf{R}\colon |\frac{3}{2} + x| \ge \frac{3}{2}\}. \end{aligned}$$

6.10 Exercise. Sketch the graphs of the functions from **R** to **R** defined by the following equations:

$$\begin{aligned} f_1(x) &= |x|, \\ f_2(x) &= |x-2|, \\ f_3(x) &= |x| - |x-2|, \\ f_4(x) &= |x| + |x-2|, \\ f_5(x) &= x^2 - 1, \\ f_6(x) &= |x^2 - 1|, \\ f_7(x) &= |x^2 - 1|^2. \end{aligned}$$

(No explanations are expected for this problem.)

6.11 Exercise. Let $f_1, \cdots, f_7$ be the functions described in the previous exercise. By looking at the graphs, express each of the following six sets in terms of intervals.

$$\begin{aligned} S_1 &= \{x \in \mathbf{R}\colon f_1(x) < 1\} \\ S_2 &= \{x \in \mathbf{R}\colon f_2(x) < 1\} \\ S_3 &= \{x \in \mathbf{R}\colon f_3(x) < 1\} \\ S_4 &= \{x \in \mathbf{R}\colon f_4(x) < 3\} \\ S_5 &= \{x \in \mathbf{R}\colon f_5(x) < 3\} \\ S_6 &= \{x \in \mathbf{R}\colon f_6(x) < 3\}. \end{aligned}$$

Let $S_7 = \{x \in \mathbf{R}\colon f_7(x) < \frac{1}{2}\}$. Represent S_7 graphically on a number line.

Remark: The notation $|x|$ for absolute value of x was introduced by Weierstrass in 1841 [15][Vol 2,page 123]. It was first introduced in connection with complex numbers. It is surprising that analysis advanced so far without introducing a special notation for this very important function.

6.2 Approximation

6.12 Definition (b approximates a.) Let ϵ be a positive number, and let a and b be arbitrary numbers. I will say that b *approximates* a *with an error smaller than* ϵ if and only if

$$|b - a| < \epsilon.$$

Remark: If b approximates a with an error smaller than ϵ, then a approximates b with an error smaller than ϵ, since $|a-b| = |b-a|$.

6.13 Definition (Approximation to n decimals.) Let $n \in \mathbf{Z}^+$, and let a, b be real numbers. I will say that b *approximates a with n decimal accuracy* if and only if b approximates a with an error smaller than $\frac{1}{2} \cdot 10^{-n}$; i.e., if and only if

$$|b-a| < \frac{1}{2}10^{-n}.$$

6.14 Notation. If I write three dots $(\cdots)$ at the end of a number written in decimal notation, I assume that all of the digits before the three dots are correct. Thus since $\pi = 3.141592653589\cdots$, I have $\pi = 3.1415\cdots$, and $\pi = 3.1416$ with 4 decimal accuracy.

6.15 Example.

$$\pi = 3.141592653589793\cdots$$

and

$$\frac{22}{7} = 3.142857142857142\cdots.$$

Hence

$$3.1415 < \pi < \frac{22}{7} < 3.1429,$$

and

$$\left|\frac{22}{7} - \pi\right| < 3.1429 - 3.1415 = .0014 < .005 = \frac{1}{2} \cdot 10^{-2}.$$

Hence $\frac{22}{7}$ approximates π with an error smaller than .0014, and $\frac{22}{7}$ approximates π with 2 decimal accuracy.

6.16 Example. We see that

.49 approximates .494999 with 2 decimal accuracy,

and

.50 approximates .495001 with 2 decimal accuracy,

but there is no two digit decimal that approximates .495000 with 2 decimal accuracy.

6.17 Example. Since

$$|.49996 - .5| = .00004 < .00005 = \frac{1}{2} \cdot 10^{-4},$$

we see that .5 approximates .49996 with 4 decimal accuracy, even though the two numbers have no decimal digits in common. Since

$$|.49996 - .4999| = .00006 > \frac{1}{2} \cdot 10^{-4},$$

we see that .4999 does not approximate .49996 with 4 decimal accuracy, even though the two numbers have four decimal digits in common.

6.18 Theorem (Strong approximation theorem.) *Let a and b be real numbers. Suppose that for every positive number ϵ, b approximates a with an error smaller than ϵ. Then $b = a$.*

Proof: Suppose that b approximates a with an error smaller than ϵ for every positive number ϵ. Then

$$|b - a| < \epsilon \text{ for every } \epsilon \text{ in } \mathbf{R}^+.$$

Hence

$$|b - a| \neq \epsilon \text{ for every } \epsilon \text{ in } \mathbf{R}^+,$$

i.e., $|b - a| \notin \mathbf{R}^+$. But $|b - a| \in \mathbf{R}_{\geq 0}$, so it follows that $|b - a| = 0$, and consequently $b - a = 0$; i.e., $b = a$. $\|$

6.3 Convergence of Sequences

6.19 Definition ($\{a_n\}$ converges to L.)

Let $\{a_n\}$ be a sequence of real numbers, and let L be a real number. We say that $\{a_n\}$ *converges to* L if for every positive number ϵ there is a number $N(\epsilon)$ in $\mathbf{Z}^+$ such that all of the numbers a_n for which $n \geq N(\epsilon)$ approximate L with an error smaller than ϵ. We denote the fact that $\{a_n\}$ converges to L by the notation

$$\{a_n\} \to L.$$

Thus "$\{a_n\} \to L$" means:

For every $\epsilon \in \mathbf{R}^+$ there is a number $N(\epsilon)$ in $\mathbf{Z}^+$ such that

$$|a_n - L| < \epsilon \text{ for all } n \text{ in } \mathbf{Z}^+ \text{ with } n \geq N(\epsilon).$$

Since

$$|a_n - L| = |(a_n - L) - 0| = \Big||a_n - L| - 0\Big|,$$

it follows immediately from the definition of convergence that

$$(\{a_n\} \to L) \iff (\{a_n - L\} \to 0) \iff (|a_n - L| \to 0).$$

We will make frequent use of these equivalences.

6.20 Example. If $a \in \mathbf{R}^+$ then

$$\left\{a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right)\right\} \to a^3.$$

Proof: Let ϵ be a generic element of $\mathbf{R}^+$. I must find a number $N(\epsilon)$ such that

$$\left|a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right) - a^3\right| < \epsilon \tag{6.21}$$

whenever $n \geq N(\epsilon)$. Well, for all n in $\mathbf{Z}^+$

$$\begin{aligned}
\left|a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right) - a^3\right| &= \left|a^3\left(1+\frac{3}{2n}+\frac{1}{2n^2}\right) - a^3\right| \\
&= \left|a^3\left(\frac{3}{2n}+\frac{1}{2n^2}\right)\right| = a^3\left(\frac{3}{2n}+\frac{1}{2n^2}\right) \\
&\leq a^3\left(\frac{3}{2n}+\frac{1}{2n}\right) = \frac{2a^3}{n}.
\end{aligned} \tag{6.22}$$

Now for every n in $\mathbf{Z}^+$ we have

$$\left(\frac{2a^3}{n} < \epsilon\right) \iff \left(\frac{2a^3}{\epsilon} < n\right),$$

and by the Archimedean property of $\mathbf{R}$ there is some integer $N(\epsilon)$ such that $\frac{2a^3}{\epsilon} < N(\epsilon)$. For all $n \geq N(\epsilon)$ we have

$$(n \geq N(\epsilon)) \implies \left(\frac{2a^3}{\epsilon} < N(\epsilon) \leq n\right) \implies \left(\left(\frac{2a^3}{n}\right) < \epsilon\right),$$

so by (6.22)

$$(n \geq N(\epsilon)) \implies \left|a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right)-a^3\right| \leq \frac{2a^3}{n} < \epsilon.$$

Hence by the definition of convergence we have

$$\left\{a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right)\right\} \to a^3. \parallel \tag{6.23}$$

A very similar argument can be used to show that

$$\left\{a^3\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)\right\} \to a^3. \tag{6.24}$$

6.25 Example. In the eighteenth century the rather complicated argument just given would have been stated as

> If n is infinitely large, then $a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right) = a^3$.

The first calculus text book (written by Guillaume François de l'Hôpital and published in 1696) sets forth the postulate

> Grant that two quantities, whose difference is an infinitely small quantity, may be taken (or used) indifferently for each other: or (which is the same thing) that a quantity which is increased or decreased only by an infinitely small quantity, may be considered as remaining the same[35, page 314].

If n is infinite, then $\frac{1}{n}$ is infinitely small, so $\left(1+\frac{1}{n}\right) = 1$, and similarly $\left(1+\frac{1}{2n}\right) = 1$. Hence

$$a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right) = a^3 \cdot 1 \cdot 1 = a^3.$$

There were numerous objections to this sort of reasoning. Even though $\left(1+\frac{1}{n}\right) = 1$, we do not have $\left(1+\frac{1}{n}\right) - 1 = 0$, since

$$\frac{\left(1+\frac{1}{n}\right)-1}{\frac{1}{n}} = 1.$$

It took many mathematicians working over hundreds of years to come up with our definition of convergence.

6.26 Theorem (Uniqueness theorem for convergence.) *Let $\{a_n\}$ be a sequence of real numbers, and let a, b be real numbers. Suppose*

$$\{a_n\} \to a \text{ and } \{a_n\} \to b.$$

Then $a = b$.

Proof: Suppose $\{a_n\} \to a$ and $\{a_n\} \to b$. By the triangle inequality

$$|a - b| = |(a - a_n) - (b - a_n)| \leq |a - a_n| + |b - a_n|. \tag{6.27}$$

Let ϵ be a generic element of $\mathbf{R}^+$. Then $\frac{\epsilon}{2}$ is also in $\mathbf{R}^+$. Since $\{a_n\} \to a$, there is a number $N(\frac{\epsilon}{2})$ in $\mathbf{Z}^+$ such that

$$|a - a_n| < \frac{\epsilon}{2} \text{ for all } n \geq N(\tfrac{\epsilon}{2}). \tag{6.28}$$

Since $\{a_n\} \to b$ there is a number $M(\frac{\epsilon}{2})$ in $\mathbf{Z}^+$ such that

$$|b - a_n| < \frac{\epsilon}{2} \text{ for all } n \geq M(\tfrac{\epsilon}{2}). \tag{6.29}$$

Let $P(\epsilon)$ be the larger of $N(\frac{\epsilon}{2})$ and $M(\frac{\epsilon}{2})$. If n is a positive integer and $n \geq P(\epsilon)$ then by (6.27), (6.28), and (6.29), we have

$$|a - b| \leq |a - a_n| + |b - a_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since this holds for all ϵ in $\mathbf{R}^+$, we have $a = b$. |||

6.30 Definition (Limit of a sequence.) Let $\{a_n\}$ be a sequence of real numbers. If there is a number a such that $\{a_n\} \to a$, we write $\lim\{a_n\} = a$. The uniqueness theorem for convergence shows that this definition makes sense. If $\lim\{a_n\} = a$, we say *a is the limit of the sequence $\{a_n\}$.*

6.31 Definition (Convergent and divergent sequence.) Let $\{a_n\}$ be a sequence of real numbers. If there is a number a such that $\{a_n\} \to a$, we say that $\{a_n\}$ is a *convergent sequence.* If there is no such number a, we say that $\{a_n\}$ is a *divergent sequence.*

6.32 Example. It follows from example 6.20 that

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

for all a in $\mathbf{R}^+$. Hence $\left\{a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right)\right\}$ is a convergent sequence for each a in $\mathbf{R}^+$.

The sequence $\{n\}$ is a divergent sequence. To see this, suppose there were a number a such that $\{n\} \to a$.

Then we can find a number $N(\frac{1}{3})$ such that

$$|n-a| < \frac{1}{3} \text{ for all } n \geq N(\tfrac{1}{3}).$$

In particular

$$\left|N(\tfrac{1}{3}) - a\right| < \frac{1}{3} \text{ and } \left|\left(N(\tfrac{1}{3})+1\right) - a\right| < \frac{1}{3}$$

(since $N(\frac{1}{3})+1$ is an integer greater than $N(\frac{1}{3})$). Hence, by the triangle inequality

$$\begin{aligned} 1 &= |1| = \left|\left(N(\tfrac{1}{3})+1-a\right) - \left(N(\tfrac{1}{3})-a\right)\right| \\ &\leq \left|N(\tfrac{1}{3})+1-a\right| + \left|N(\tfrac{1}{3})-a\right| < \frac{1}{3}+\frac{1}{3} = \frac{2}{3} \end{aligned}$$

i.e., $1 < \dfrac{2}{3}$ which is false.

Since the assumption $\{n\} \to a$ has led to a contradiction, it is false that $\{n\} \to a$. |||

6.33 Exercise. Let $\{a_n\}$ be a sequence of real numbers, and let a be a real number. Suppose that as n gets larger and larger, a_n gets nearer and nearer to a, i.e., suppose that for all m and n in $\mathbf{Z}^+$

$$(n > m) \implies (|a_n - a| < |a_m - a|).$$

Does it follow that $\{a_n\}$ converges to a?

6.34 Exercise. For each of the sequences below, calculate the first few terms, and make a guess as to whether or not the sequence converges. In some cases you will need to use a calculator. Try to explain the basis for your guess. (If you can prove your guess is correct, do so, but in several cases the proofs involve more mathematical knowledge than you now have.)

$\{a_n\} = \{(-1)^n\}$.

$\{c_n\} = \left\{1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n+1}}{n}\right\}$.

$\{d_n\} = \left\{1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}\right\}$
This problem was solved by Leonard Euler (1707-1783)[18, pp138-139].

$\{e_n\} = \left\{1 + \frac{1}{3} + \frac{1}{3^2} + \cdots + \frac{1}{3^{n-1}}\right\}$.

$\{f_n\} = \left\{\left(1 + \frac{1}{n}\right)^n\right\}$
This problem was solved by Jacob Bernoulli (1654-1705)[8, pp94-97].

.

6.4 Properties of Limits.

In this section I will state some basic properties of limits. All of the statements listed here as assumptions are, in fact, theorems that can be proved from the definition of limits. I am omitting the proofs because of lack of time, and because the results are so plausible that you will probably believe them without a proof.

6.35 Definition (Constant sequence.) If r is a real number then the sequence $\{r\}$ all of whose terms are equal to r is called a *constant sequence*

$$\{r\} = \{r, r, r, \cdots\}.$$

It is an immediate consequence of the definition of convergence that

$$\{r\} \to r$$

for every real number r. (If $r_n = r$ for all n in $\mathbf{Z}^+$ then $|r_n - r| = 0 < \epsilon$ for all ϵ in $\mathbf{R}^+$ so r_n approximates r with an error smaller than ϵ for all $n \geq 1$. $\|\!|$.)

We have just proved

6.36 Theorem (Constant sequence rule.) *If $\{r\}$ denotes a constant sequence of real numbers, then*

$$\lim\{r\} = r.$$

6.37 Theorem (Null sequence rule.) *Let α be a positive rational number. Then*

$$\lim\left\{\frac{1}{n^\alpha}\right\} = 0.$$

Proof: Let α be a positive rational number, and let ϵ be a generic positive number. By the monotonicity of powers (see (C.95) in appendix C), we have

$$\begin{aligned}\frac{1}{n^\alpha} < \epsilon \quad &\iff \quad \left(\frac{1}{n^\alpha}\right)^{\frac{1}{\alpha}} < (\epsilon)^{\frac{1}{\alpha}} \iff \frac{1}{n} < \epsilon^{\left(\frac{1}{\alpha}\right)} \\ &\iff \quad n > \frac{1}{\epsilon^{\frac{1}{\alpha}}} = \epsilon^{-\frac{1}{\alpha}}.\end{aligned}$$

By the Archimedean property for $\mathbf{R}$ there is an integer $N(\epsilon)$ in $\mathbf{Z}^+$ such that

$$N(\epsilon) > \epsilon^{-\frac{1}{\alpha}}.$$

Then for all n in $\mathbf{Z}^+$

$$n \geq N(\epsilon) \implies n \geq \epsilon^{-1/\alpha} \implies \frac{1}{n^\alpha} < \epsilon \implies \left|\frac{1}{n^\alpha} - 0\right| < \epsilon.$$

Thus $\lim\left\{\frac{1}{n^\alpha}\right\} = 0.$ |||

6.38 Assumption (Sum rule for sequences.) *Let $\{a_n\}$ and $\{b_n\}$ be convergent sequences of real numbers. Then*

$$\lim\{a_n + b_n\} = \lim\{a_n\} + \lim\{b_n\}$$

and

$$\lim\{a_n - b_n\} = \lim\{a_n\} - \lim\{b_n\}.$$

The sum rule is actually easy to prove, but I will not prove it. (You can probably supply a proof for it.)

Notice the hypothesis that $\{a_n\}$ and $\{b_n\}$ are *convergent* sequences. It is not true in general that

$$\lim\{a_n + b_n\} = \lim\{a_n\} + \lim\{b_n\}.$$

For example, the statement

$$\lim\{(-1)^n + (-1)^{n+1}\} = \lim\{(-1)^n\} + \lim\{(-1)^{n+1}\}$$

is false, since

$$\lim\{(-1)^n + (-1)^{n+1}\} = \lim\{0\} = 0$$

but neither of the limits $\lim\{(-1)^n\}$ or $\lim\{(-1)^{n+1}\}$ exist.

6.39 Assumption (Product rule for sequences.) *Let $\{a_n\}$ and $\{b_n\}$ be convergent sequences. Then*

$$\lim\{a_n \cdot b_n\} = \lim\{a_n\} \cdot \lim\{b_n\}.$$

An important special case of the product rule occurs when one of the sequences is constant: If a is a real number, and $\{b_n\}$ is a convergent sequence, then

$$\lim\{ab_n\} = a \lim\{b_n\}.$$

The intuitive content of the product rule is that if a_n approximates a very well, and b_n approximates b very well, then $a_n b_n$ approximates ab very well. It is somewhat tricky to prove this for a reason that is illustrated by the following example.

According to Maple,

$$\sqrt{99999999} = 9999.99994999999987499\cdots$$

so 9999.9999 approximates $\sqrt{99999999}$ with 4 decimal accuracy. Let

$$a = b = 9999.9999,$$

and let

$$A = B = \sqrt{99999999}.$$

Then a approximates A with 4 decimal accuracy and b approximates B with 4 decimal accuracy. But

$$AB = 99999999$$

and

$$ab = 99999998.00000001$$

so ab does not approximate AB with an accuracy of even one decimal.

6.40 Assumption (Quotient rule for sequences.) *Let $\{a_n\}$ and $\{b_n\}$ be convergent real sequences such that $b_n \neq 0$ for all n in $\mathbf{Z}^+$ and $\lim\{b_n\} \neq 0$. Then*

$$\lim\left\{\frac{a_n}{b_n}\right\} = \frac{\lim\{a_n\}}{\lim\{b_n\}}.$$

The hypotheses here are to be expected. If some term b_n were zero, then $\left\{\frac{a_n}{b_n}\right\}$ would not be a sequence, and if $\lim\{b_n\}$ were zero, then $\frac{\lim\{a_n\}}{\lim\{b_n\}}$ would not be defined.

6.41 Assumption (Inequality rule for sequences.) *Let $\{a_n\}$ and $\{b_n\}$ be convergent sequences. Suppose there is an integer N in $\mathbf{Z}^+$ such that*

$$a_n \leq b_n \text{ for all } n \text{ in } \mathbf{Z}_{\geq N}.$$

Then

$$\lim\{a_n\} \leq \lim\{b_n\}.$$

The most common use of this rule is in situations where

$$0 \leq b_n \text{ for all } n$$

and we conclude that

$$0 \leq \lim\{b_n\}.$$

6.42 Assumption (Squeezing rule for sequences.) *Let $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ be three real sequences. Suppose there is an integer N in $\mathbf{Z}^+$ such that*

$$a_n \leq b_n \leq c_n \text{ for all } n \in \mathbf{Z}_{\geq N}. \tag{6.43}$$

Suppose further, that $\{a_n\}$ and $\{c_n\}$ both converge to the same limit L. Then $\{b_n\}$ also converges to L.

If we knew that the middle sequence, $\{b_n\}$ in the squeezing rule was convergent, then we would be able to prove the squeezing rule from the inequality rule, since if all three sequences $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ converge, then it follows from (6.43) that

$$\lim\{a_n\} \leq \lim\{b_n\} \leq \lim\{c_n\},$$

i.e.

$$L \leq \lim\{b_n\} \leq L$$

and hence $\lim\{b_n\} = L$. The power of the squeezing rule is that it allows us to conclude that a limit exists.

6.44 Definition (Translate of a sequence.) Let $\{a_n\}$ be a real sequence, and let $p \in \mathbf{Z}^+$. The sequence $\{a_{n+p}\}$ is called a *translate* of $\{a_n\}$.

6.45 Example. If

$$\{a_n\} = \left\{\frac{1}{n^2}\right\} = \left\{1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \cdots\right\}$$

then

$$\{a_{n+2}\} = \left\{\frac{1}{(n+2)^2}\right\} = \left\{\frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \cdots\right\}.$$

If

$$\{b_n\} = \{(-1)^n\}$$

then

$$\{b_{n+2}\} = \{(-1)^{n+2}\} = \{(-1)^n\} = \{b_n\}.$$

6.46 Theorem (Translation rule for sequences.) *Let $\{a_n\}$ be a convergent sequence of real numbers, and let p be a positive integer. Then $\{a_{n+p}\}$ is convergent and*

$$\lim\{a_{n+p}\} = \lim\{a_n\}.$$

Proof: Suppose $\lim\{a_n\} = a$, and let ϵ be a generic element in $\mathbf{R}^+$. Then we can find an integer $N(\epsilon)$ in $\mathbf{Z}^+$ such that

$$|a_n - a| < \epsilon \text{ for all } n \text{ in } \mathbf{Z}^+ \text{ with } n \geq N(\epsilon).$$

If $n \geq N(\epsilon)$ then $n + p \geq N(\epsilon) + p \geq N(\epsilon)$ so

$$|a_{n+p} - a| < \epsilon.$$

This shows that $\lim\{a_{n+p}\} = a = \lim\{a_n\}$. $\|$

6.47 Example. The sequence

$$\{a_n\} = \left\{\frac{1}{n+4}\right\} = \left\{\frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \cdots\right\}$$

is a translate of the sequence $\left\{\frac{1}{n}\right\}$. Since $\lim\left\{\frac{1}{n}\right\} = 0$ it follows from the translation theorem that $\lim\left\{\frac{1}{n+4}\right\} = 0$ also.

6.48 Theorem (nth root rule for sequences.) *Let a be a positive number then*

$$\lim\left\{a^{\frac{1}{n}}\right\} = 1.$$

Proof: Case 1: Suppose $a = 1$. Then

$$\lim\left\{a^{\frac{1}{n}}\right\} = \lim\{1\} = 1.$$

Case 2: Suppose $a > 1$, so that $a^{\frac{1}{n}} > 1$ for all $n \in \mathbf{Z}^+$. Let ϵ be a generic positive number, and let n be a generic element of $\mathbf{Z}^+$. Since ln is strictly increasing on $\mathbf{R}^+$ we have

$$\begin{aligned}\left(a^{\frac{1}{n}} - 1 < \epsilon\right) &\iff \left(a^{\frac{1}{n}} < 1 + \epsilon\right) \iff \left(\ln(a^{\frac{1}{n}}) < \ln(1+\epsilon)\right) \\ &\iff \frac{1}{n}\ln(a) < \ln(1+\epsilon) \\ &\iff \frac{\ln(a)}{\ln(1+\epsilon)} < n. \qquad (6.49)\end{aligned}$$

(In the last step I used the fact that $\ln(1+\epsilon) > 0$ if $\epsilon > 0$.) By the Archimedean property for $\mathbf{R}$ there is an integer $N(\epsilon)$ in $\mathbf{Z}^+$ such that

$$\frac{\ln(a)}{\ln(1+\epsilon)} < N(\epsilon).$$

For all $n \in \mathbf{Z}^+$ we have

$$\begin{aligned} n \geq N(\epsilon) &\implies \frac{\ln(a)}{\ln(1+\epsilon)} < N(\epsilon) \leq n \\ &\implies a^{\frac{1}{n}} - 1 < \epsilon \implies \left|a^{\frac{1}{n}} - 1\right| < \epsilon.\end{aligned}$$

Hence $\lim\left\{a^{\frac{1}{n}}\right\}=1$.
Case 3: Suppose $0<a<1$. Then $a^{-1}>1$ so by Case 2, we have

$$\begin{aligned}\lim\left\{a^{\frac{1}{n}}\right\} &= \lim\left\{\frac{1}{(a^{-1})^{\frac{1}{n}}}\right\}\\ &= \frac{\lim\{1\}}{\lim\left\{(a^{-1})^{\frac{1}{n}}\right\}}=\frac{1}{1}=1.\end{aligned}$$

Thus, in all cases, we have

$$\lim\left\{a^{\frac{1}{n}}\right\}=1. \ \|$$

6.5 Illustrations of the Basic Limit Properties.

6.50 Example. In example 6.20, we used the definition of limit to show that

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

for all $a\in\mathbf{R}^+$, and claimed that a similar argument shows that

$$\lim\left\{a^3\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)\right\}=a^3 \tag{6.51}$$

We will now use the basic properties of limits to prove (6.51). By the product rule and the null sequence rule,

$$\lim\left\{\frac{1}{2n}\right\}=\lim\left\{\frac{1}{2}\cdot\frac{1}{n}\right\}=\frac{1}{2}\lim\left\{\frac{1}{n}\right\}=\frac{1}{2}\cdot 0=0.$$

Hence by the sum rule

$$\lim\left\{1-\frac{1}{2n}\right\}=\lim\{1\}-\lim\left\{\frac{1}{2n}\right\}=1-0=1.$$

By the sum rule and the null sequence rule

$$\lim\left\{1-\frac{1}{n}\right\}=\lim\{1\}-\lim\left\{\frac{1}{n}\right\}=1-0=1.$$

Hence by the product rule,

$$\begin{aligned}\lim\left\{\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{1}{2n}\right)\right\} &= \lim\left\{\left(1-\frac{1}{n}\right)\right\}\cdot\lim\left\{\left(1-\frac{1}{2n}\right)\right\}\\ &= 1\cdot 1=1.\end{aligned}$$

Now $\{a^3\}$ is a constant sequence, so by the product rule,

$$\begin{aligned}\lim\left\{a^3\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)\right\} &= a^3\cdot\lim\left\{\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{1}{2n}\right)\right\}\\ &= a^3\cdot 1 = a^3.\end{aligned}$$

6.52 Example. In the previous example, I made at least eight applications of our limit rules. However, the applications are completely mechanical so I will usually not be so careful, and in a situation like this, I will just write

$$\lim\left\{a^3\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)\right\} = a^3\cdot(1-0)\cdot\left(1-\frac{1}{2}\cdot 0\right) = a^3. \tag{6.53}$$

The argument given in equation (6.53) looks remarkably similar to the eighteenth century argument given in example 6.25.

6.54 Example. Let a be a positive number, and let

$$A(a) = \text{area}(\{(x,y)\in\mathbf{R}^2\colon 0\le x\le a \text{ and } 0\le y\le x^2\}).$$

In (2.13), we showed that

$$\frac{a^3}{3}\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right) \le\ A(a) \le \frac{a^3}{3}\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right), \tag{6.55}$$

for all $n\in\mathbf{Z}^+$, and claimed that these inequalities show that $A(a)=\dfrac{a^3}{3}$. Now I want to examine the claim more closely.

In example 6.20 we proved that

$$\lim\left\{a^3\left(1+\frac{1}{n}\right)\left(1+\frac{1}{2n}\right)\right\} = a^3,$$

and in example 6.50 we proved that

$$\lim\left\{a^3\left(1-\frac{1}{n}\right)\left(1-\frac{1}{2n}\right)\right\} = a^3$$

By applying the squeezing rule to equation 6.55, we see that

$$\lim\{A(a)\} = a^3,$$

i.e.

$$A(a) = \frac{a^3}{3}.\ \|$$

6.56 Example. I will now consider the limit

$$\lim\left\{\frac{n^2-2n}{n^2+3n}\right\}.$$

Here I cannot apply the quotient rule for sequences, because the limits of the numerator and denominator do not exist. However, I notice that I can simplify my sequence:

$$\left\{\frac{n^2-2n}{n^2+3n}\right\}=\left\{\frac{n-2}{n+3}\right\}.$$

I will now use a trick. I will factor the highest power of n out of the numerator and denominator:

$$\left\{\frac{n-2}{n-3}\right\}=\left\{\frac{n\left(1-\frac{2}{n}\right)}{n\left(1-\frac{3}{n}\right)}\right\}=\left\{\frac{1-\frac{2}{n}}{1-\frac{3}{n}}\right\}.$$

It is now clear what the limit is.

$$\lim\left\{\frac{n^2-2n}{n^2+3n}\right\}=\lim\left\{\frac{1-\frac{2}{n}}{1-\frac{3}{n}}\right\}=\frac{1-2\cdot 0}{1-3\cdot 0}=1.$$

6.57 Example. I want to investigate

$$\lim\left\{\frac{n-2n^2+3}{4+6n+n^2}.\right\}$$

I'll apply the factoring trick of the previous example.

$$\left\{\frac{n-2n^2+3}{4+6n+n^2}\right\}=\left\{\frac{n^2\left(\frac{1}{n}-2+\frac{3}{n^2}\right)}{n^2\left(\frac{4}{n^2}+\frac{6}{n}+1\right)}\right\}=\left\{\frac{\frac{1}{n}-2+\frac{3}{n^2}}{\frac{4}{n^2}+\frac{6}{n}+1}\right\}$$

so

$$\begin{aligned}\lim\left\{\frac{n-2n^2+3}{4+6n+n^2}\right\} &= \lim\left\{\frac{\frac{1}{n}-2+\frac{3}{n^2}}{\frac{4}{n^2}+\frac{6}{n}+1}\right\}=\frac{0-2+3\cdot 0}{4\cdot 0+6\cdot 0+1}\\ &= -2.\end{aligned}$$

6.58 Example. I want to find

$$\lim\left\{\frac{1}{n+4}\right\}.$$

I observe that $\left\{\dfrac{1}{n+4}\right\}$ is a translate of $\left\{\dfrac{1}{n}\right\}$ so by the translation rule

$$\lim\left\{\frac{1}{n+4}\right\} = \lim\left\{\frac{1}{n}\right\} = 0.$$

I can also try to do this by my factoring trick:

$$\begin{aligned}\lim\left\{\frac{1}{n+4}\right\} &= \lim\left\{\frac{1}{n\left(1+\frac{4}{n}\right)}\right\} = \lim\left\{\frac{\frac{1}{n}}{1+\frac{4}{n}}\right\}\\ &= \frac{0}{1+4\cdot 0} = 0.\end{aligned}$$

6.59 Exercise. Find the following limits, or explain why they don't exist.

a) $\lim\left\{7+\dfrac{6}{n}+\dfrac{8}{\sqrt{n}}\right\}$

b) $\lim\left\{\dfrac{4+\frac{1}{n}}{5+\frac{1}{n}}\right\}$

c) $\lim\left\{\dfrac{3n^2+n+1}{1+3n+4n^2}\right\}$

d) $\lim\left\{\dfrac{\left(2+\frac{1}{n}\right)^2+4}{\left(2+\frac{1}{n}\right)^3+8}\right\}$

e) $\lim\left\{\dfrac{\left(2+\frac{1}{n}\right)^2-4}{\left(2+\frac{1}{n}\right)^3-8}\right\}$

f) $\lim\left\{\dfrac{8n^3+13n}{17+12n^3}\right\}$

g) $\lim\left\{\dfrac{8(n+4)^3+13(n+4)}{17+12(n+4)^3}\right\}$

h) $\lim\left\{\dfrac{n+1}{n^2+1}\right\}$.

6.60 Example. Let a be a real number greater than 1, and let

$$S_a = \{(x,y) \in \mathbf{R}^2 \colon 1 \le x \le a \text{ and } 0 \le y \le \frac{1}{x^2}\}.$$

In (2.34) we showed that

$$\frac{(1-a^{-1})}{a^{\frac{1}{n}}} \le \text{area}(S_a) \le a^{\frac{1}{n}}(1-a^{-1}) \text{ for all } n \in \mathbf{Z}^+. \tag{6.61}$$

I want to conclude from this that $\text{area}(S_a) = (1-a^{-1})$.

By the nth root rule, and the quotient and product rules, we have

$$\lim\left\{\frac{(1-a^{-1})}{a^{\frac{1}{n}}}\right\} = \frac{\lim\{1-a^{-1}\}}{\lim\{a^{\frac{1}{n}}\}} = \frac{(1-a^{-1})}{1} = (1-a^{-1}),$$

and

$$\lim\left\{a^{\frac{1}{n}}(1-a^{-1})\right\} = \lim\left\{a^{\frac{1}{n}}\right\}\lim\{(1-a^{-1})\} = 1\cdot(1-a^{-1}) = (1-a^{-1}).$$

By (6.61) and the squeezing rule, we conclude that

$$\lim\{\text{area}(S_a)\} = (1-a^{-1}),$$

i.e.

$$\text{area}(S_a) = (1-a^{-1}).$$

6.62 Example. Let the sequence $\{a_n\}$ be defined by the rules

$$\begin{aligned} a_1 &= 1, \\ a_{n+1} &= \frac{a_n^2+2}{2a_n} \text{ for } n \ge 1. \end{aligned} \tag{6.63}$$

Thus, for example

$$a_2 = \frac{1+2}{2} = \frac{3}{2}$$

and

$$a_3 = \frac{\frac{9}{4}+2}{3} = \frac{17}{12}.$$

It is clear that $a_n > 0$ for all n in $\mathbf{Z}^+$. Let $L = \lim\{a_n\}$. Then by the translation rule, $L = \lim\{a_{n+1}\}$ also. From (6.63) we have

$$2a_na_{n+1} = a_n^2 + 2 \text{ for all } n \in \mathbf{Z}_{\ge 2}.$$

Thus

$$\lim\{2a_n a_{n+1}\} = \lim\{a_n^2 + 2\},$$

i.e.

$$2 \cdot \lim\{a_n\} \lim\{a_{n+1}\} = \lim\{a_n\}^2 + \lim\{2\}.$$

Hence

$$2 \cdot L \cdot L = L^2 + 2.$$

Thus $L^2 = 2$, and it follows that $L = \sqrt{2}$ or $L = -\sqrt{2}$. But we noticed above that $a_n > 0$ for all n in $\mathbf{Z}^+$, and hence by the inequality rule for sequences, $L \geq 0$. Hence we conclude that $L = \sqrt{2}$, i.e.,

$$\lim\{a_n\} = \sqrt{2}. \tag{6.64}$$

(Actually there is an error in the reasoning here, which you should try to find, but the conclusion (6.64) is in fact correct. After you have done exercise 6.68, the error should become apparent.)

6.65 Exercise. Use a calculator to find the first six terms of the sequence (6.63). Do all calculations using all the accuracy your calculator allows, and write down the results to all the accuracy you can get. Compare your answers with $\sqrt{2}$ (as given by your calculator) and for each term note how many decimal places accuracy you have.

6.66 Example. Let $\{b_n\}$ be the sequence defined by the rules

$$\begin{aligned} b_1 &= 1, \\ b_2 &= 1, \\ b_n &= \frac{1 + b_{n-1}}{b_{n-2}} \quad \text{for} \quad n > 2. \end{aligned} \tag{6.67}$$

Thus, for example

$$b_3 = \frac{1+1}{1} = 2$$

and

$$b_4 = \frac{1+2}{1} = 3.$$

Notice that $b_n > 0$ for all n. Let

$$L = \lim\{b_n\}.$$

By the translation rule

$$L = \lim\{b_{n+1}\} \text{ and } L = \lim\{b_{n+2}\}.$$

By (6.67) (with n replaced by $n+2$), we have

$$b_n b_{n+2} = 1 + b_{n+1} \text{ for all } n \text{ in } \mathbf{Z}^+.$$

Hence

$$\begin{aligned} L^2 &= \lim\{b_n\} \cdot \lim\{b_{n+2}\} \\ &= \lim\{b_n b_{n+2}\} \\ &= \lim\{1 + b_{n+1}\} = 1 + L. \end{aligned}$$

Thus

$$L^2 - L - 1 = 0.$$

By the quadratic formula

$$L = \frac{1+\sqrt{5}}{2} \text{ or } L = \frac{1-\sqrt{5}}{2}.$$

Since $b_n > 0$ for all n, we have $L \geq 0$, so we have

$$L = \frac{1+\sqrt{5}}{2}.$$

(This example has the same error as the previous one.)

6.68 Exercise. Repeat exercise 6.65 using the sequence $\{b_n\}$ described in (6.67) in place of the sequence $\{a_n\}$, and $\left(\frac{1+\sqrt{5}}{2}\right)$ in place of $\sqrt{2}$. After doing this problem, you should be able to point out the error in examples (6.62) and (6.66). (This example is rather surprising. I took it from [14, page 55, exercise 20].)

6.69 Exercise. For each of the statements below: if the statement is false, give a counterexample; if the statement is true, then justify it by means of limit rules we have discussed.

a) Let $\{a_n\}$ be a convergent sequence of real numbers. If $a_n > 0$ for all n in $\mathbf{Z}^+$, then $\lim\{a_n\} > 0$.

b) Let $\{a_n\}$ and $\{b_n\}$ be real sequences. If $\lim\{a_n\} = 0$, then $\lim\{a_n b_n\} = 0$.

c) Let $\{a_n\}$ be a real sequence. If $\lim\{a_n^2\} = 1$ then either $\lim\{a_n\} = 1$ or $\lim\{a_n\} = -1$.

d) Let $\{a_n\}$ and $\{b_n\}$ be real sequences. If $\lim\{a_n b_n\} = 0$, then either $\lim\{a_n\} = 0$ or $\lim\{b_n\} = 0$.

6.70 Exercise. Let a and r be positive numbers and let

$$S_0^a[t^r] = \{(x, y) \in \mathbf{R}^2 \colon 0 \le x \le a \text{ and } 0 \le y \le x^r\}.$$

In (2.4) we showed that

$$\frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + (n-1)^r) \le \alpha(S_0^a[t^r]) \le \frac{a^{r+1}}{n^{r+1}}(1^r + 2^r + \cdots + n^r).$$

Use this result, together with Bernoulli's power sums listed on page 27 to find the area of $S_0^a[t^3]$.

6.71 Theorem (nth power theorem.) *Let r be a real number such that $|r| < 1$. Then* $\lim\{r^n\} = 0$.

Proof: Let $L = \lim\{r^{n-1}\}$. Now $\{r^n\}$ is a translate of $\{r^{n-1}\}$, so by the translation theorem

$$\begin{aligned} L &= \lim\{r^{n-1}\} = \lim\{r^n\} = \lim\{r \cdot r^{n-1}\} \\ &= \lim\{r\} \lim\{r^{n-1}\} = rL \end{aligned}$$

so we have $L - rL = 0$ or

$$L(1 - r) = 0.$$

We assumed that $|r| < 1$, so $1 - r \ne 0$, and hence it follows that $L = 0$. |||

The proof just given is not valid. In fact, the argument shows that $\lim\{r^n\} = 0$ whenever $r \ne 1$, and this is certainly wrong when $r = 2$. The error comes in the first sentence, "Let $L = \lim\{r^{n-1}\}$". The argument works if the sequence $\{r^{n-1}\}$ or $\{r^n\}$ converges. We will now give a second (correct) proof of theorem 6.71.

Second Proof: Let r be a real number with $|r| < 1$. If $r = 0$, then $\{r^n\} = \{0\}$ is a constant sequence, and $\lim\{r^n\} = \lim\{0\} = 0$. Hence the theorem holds when $r = 0$, and we may assume that $r \neq 0$. Let ϵ be a generic positive number, If $n \in \mathbf{Z}^+$ we have

$$(|r^n - 0| < \epsilon) \iff (|r|^n < \epsilon) \iff (n\ln(|r|) < \ln(\epsilon)).$$

Now since $|r| < 1$, we know that $\ln(|r|) < 0$ and hence

$$(n\ln(|r|) < \ln(\epsilon)) \iff \left(n > \frac{\ln(\epsilon)}{\ln(|r|)}\right).$$

By the Archimedean property, there is some positive integer $N(\epsilon)$ such that $N(\epsilon) > \dfrac{\ln(\epsilon)}{\ln(|r|)}$. Then for all n in $\mathbf{Z}^+$

$$(n \geq N(\epsilon)) \implies \left(n > \frac{\ln(\epsilon)}{\ln(|r|)}\right) \implies (|r^n - 0| < \epsilon).$$

Hence $\lim\{r^n\} = 0$. |||

6.72 Exercise. Why was it necessary to make $r = 0$ a special case in the Second Proof above?

6.6 Geometric Series

6.73 Theorem (Geometric series) *Let r be a real number such that $|r| < 1$. Then*

$$\{\sum_{i=1}^{n} r^{i-1}\} \to \frac{1}{1-r} \tag{6.74}$$

Equation (6.74) is often written in the form

$$\sum_{i=1}^{\infty} r^{i-1} = \frac{1}{1-r} \quad \text{or} \quad \sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$$

Proof: Let r be a real number such that $|r| < 1$, and for all $n \in \mathbf{Z}^+$ let

$$f(n) = \sum_{i=1}^{n} r^{i-1}.$$

Then by theorem 2.22 we have

$$f(n) = \frac{1-r^n}{1-r},$$

and hence

$$\begin{aligned}\lim\{f(n)\} &= \lim\left\{\frac{1-r^n}{1-r}\right\}\\ &= \frac{1}{1-r}\lim\{(1-r^n)\}\\ &= \frac{1}{1-r}(1-\lim\{r^n\})\end{aligned}$$

Hence by the nth power theorem

$$\lim\{f(n)\} = \frac{1}{1-r}(1-0) = \frac{1}{1-r}. \;\|$$

6.75 Exercise. Find the error in the following argument. Let R be a real number with $R \neq 1$, and for n in $\mathbf{Z}^+$, let

$$a_n = 1 + R + R^2 + \cdots + R^{n-1}.$$

Let $L = \lim\{a_n\}$. Then, by the translation rule

$$\begin{aligned}L = \lim\{a_{n+1}\} &= \lim\{1 + R + \cdots + R^n\}\\ &= \lim\{1 + R(1 + \cdots + R^{n-1})\} = \lim\{1 + Ra_n\}.\end{aligned}$$

Thus by the sum rule and product rule,

$$\begin{aligned}L &= \lim\{1\} + \lim\{Ra_n\}\\ &= 1 + R\lim\{a_n\} = 1 + RL.\end{aligned}$$

Now

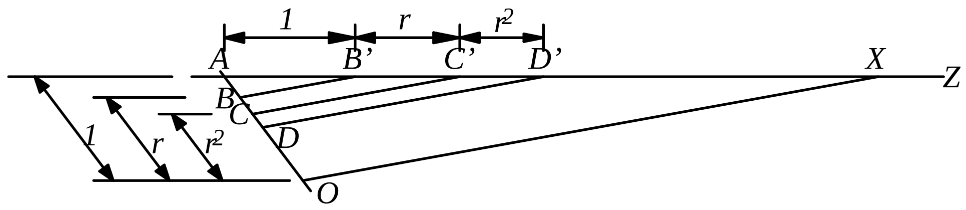
$$L = 1 + RL \implies L(1-R) = 1 \implies L = \frac{1}{1-R}.$$

Hence we have shown that

$$\lim\{1 + R + R^2 + \cdots + R^{n-1}\} = \frac{1}{1-R}$$

for all $R \in \mathbf{R} \setminus \{1\}$. (This sort of argument, *and the conclusion* were regarded as correct in the eighteenth century. At that time the argument perhaps *was* correct, because the definitions in use were not the same as ours.)

"The clearest early account of the summation of geometric series"[6, page 136] was given by Grégoire de Saint-Vincent in 1647. Grégoire's argument is roughly as follows:



On the line AZ mark off points B', C', D' etc. such that

$$AB' = 1, \;\; B'C' = r, \;\; C'D' = r^2, \;\; D'E' = r^3 \cdots$$

On a different line through A mark off points O, B, C, D etc. such that

$$OA = 1, \;\; OB = r, \;\; OC = r^2, \;\; OD = r^3 \cdots$$

Then

$$\begin{aligned} \frac{AB'}{AB} &= \frac{1}{1-r}. \\ \frac{B'C'}{BC} &= \frac{r}{r-r^2} = \frac{1}{1-r}. \\ \frac{C'D'}{CD} &= \frac{r^2}{r^2-r^3} = \frac{1}{1-r}. \\ & \textit{etc.} \end{aligned}$$

Now I use the fact that

$$\frac{a}{b} = \frac{c}{d} \Longrightarrow \frac{a+c}{b+d} = \frac{a}{b}, \tag{6.76}$$

(see exercise 6.78), to say that

$$\begin{aligned} \frac{AC'}{AC} &= \frac{AB'+B'C'}{AB+BC} = \frac{AB'}{AB} = \frac{1}{1-r} \\ \frac{AD'}{AD} &= \frac{AC'+C'D'}{AC+CD} = \frac{AC'}{AC} = \frac{1}{1-r} \\ \frac{AE'}{AE} &= \frac{AD'+D'E'}{AD+DE} = \frac{AD'}{AD} = \frac{1}{1-r} \\ & \textit{etc.} \end{aligned}$$

It follows that the triangles BAB', CAC', DAD', etc. are all mutually similar, so the lines BB', CC', DD' etc. are all parallel. Draw a line through O parallel to BB' and intersecting AZ at X. I claim that

$$AB' + B'C' + C'D' + D'E' + etc. = AX. \tag{6.77}$$

It is clear that any finite sum is smaller that AX, and by taking enough terms in the sequence $A, B, \cdots N$ we can make ON arbitrarily small. Then XN' is arbitrarily small, i.e. the finite sums AN' can be made as close to AX as we please. By similar triangles,

$$\frac{1}{1-r} = \frac{AB'}{AB} = \frac{AX}{AO} = \frac{AX}{1}$$

so, equation (6.77) says

$$1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r}.$$

6.78 Exercise. Prove the assertion (6.76).

6.79 Exercise.

a) Find $\lim \left\{1 + \left(\frac{9}{10}\right) + \left(\frac{9}{10}\right)^2 + \left(\frac{9}{10}\right)^3 + \cdots + \left(\frac{9}{10}\right)^{n-1}\right\}$.

b) Find $\lim \left\{1 - \left(\frac{9}{10}\right) + \left(\frac{9}{10}\right)^2 - \left(\frac{9}{10}\right)^3 + \cdots + \left(-\frac{9}{10}\right)^{n-1}\right\}$.

c) For each n in $\mathbf{Z}^+$ let

$$a_n = \sum_{j=0}^{\infty} \left(-\frac{n}{n+1}\right)^j,$$

(in part (b) you calculated a_9). Find a formula for a_n, and then find $\lim\{a_n\}$.

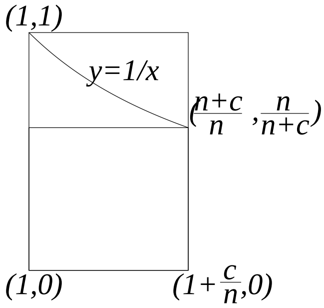
d) Show that

$$\lim \left\{\sum_{j=0}^{\infty} \left(-\frac{n}{n+1}\right)^j\right\} \neq \sum_{j=0}^{\infty} \lim \left\{\left(-\frac{n}{n+1}\right)^j\right\} \tag{6.80}$$

(Thus it is not necessarily true that the limit of an infinite sum is the infinite sum of the limits. The left side of (6.80) was calculated in part c. The right side is $\sum_{j=0}^{\infty} b_j$, where $b_j = \lim\left\{\left(-\frac{n}{n+1}\right)^j\right\}$ depends on j, but not on n.)

6.7 Calculation of e

6.81 Example. We will calculate $\lim\left\{n\ln\left(1+\frac{c}{n}\right)\right\}$, where c is a positive number. Let $f(x)=\frac{1}{x}$. Then (see the figure)



$$B(1,1+\frac{c}{n}:0,\frac{n}{n+c})\subset S_1^{1+\frac{c}{n}}f\subset B(1,1+\frac{c}{n}:0,1)$$

and hence

$$\text{area}(B(1,1+\frac{c}{n}:0,\frac{n}{n+c}))\le \text{area}(S_1^{1+\frac{c}{n}}f)\le \text{area}(B(1,1+\frac{c}{n}:0,1)).$$

Thus

$$\frac{c}{n}\cdot\frac{n}{n+c}\le \ln(1+\frac{c}{n})\le\frac{c}{n},$$

i.e.

$$\frac{cn}{n+c}\le n\ln(1+\frac{c}{n})\le c. \tag{6.82}$$

Since

$$\lim\left\{\frac{cn}{n+c}\right\}=\lim\left\{\frac{c}{1+\frac{c}{n}}\right\}=c,$$

it follows from the squeezing rule that

$$\lim\left\{n\ln\left(1+\frac{c}{n}\right)\right\}=c. \tag{6.83}$$

Notice that in this example the squeezing rule has allowed us to prove the existence of a limit whose existence was not obvious.

6.84 Example. We will show that for all $c \in \mathbf{Q}^+$

$$\lim\left\{\left(1+\frac{c}{n}\right)^n\right\} = e^c. \tag{6.85}$$

Let $c \in \mathbf{Q}^+$, and let $n \in \mathbf{Z}^+$. Let

$$a_n = \left(1+\frac{c}{n}\right)^n.$$

By (6.82), we have

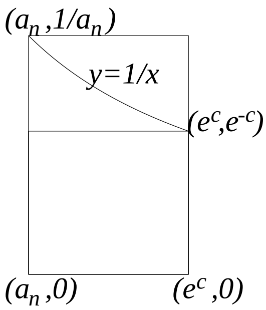
$$\ln(a_n) \leq c = c\ln(e) = \ln(e^c),$$

so

$$a_n \leq e^c \text{ for all } n \in \mathbf{Z}^+.$$

It follows from (6.83) that

$$\lim\{\ln(a_n)\} = c, \text{ or } \lim\{c - \ln(a_n)\} = 0. \tag{6.86}$$



From the picture, we see that

$$0 \leq B(a_n, e^c : 0, e^{-c}) \leq A_{a_n}^{e^c}\left[\frac{1}{t}\right],$$

i.e.

$$0 \leq e^{-c}(e^c - a_n) \leq \ln(e^c) - \ln(a_n) = c - \ln(a_n).$$

Hence

$$0 \leq e^c - a_n \leq e^c(c - \ln(a_n)), \text{ for all } n \in \mathbf{Z}^+.$$

By (6.86), we have

$$\lim\{e^c(c - \ln(a_n))\} = 0,$$

so by the squeezing rule, $\lim\{e^c - a_n\} = 0$, i.e.

$$\lim\{a_n\} = e^c.$$

This completes the proof of (6.85).

The reason we assumed c to be positive in the previous example was to guarantee that $(1+\frac{c}{n})$ has a logarithm. We could extend this proof to work for arbitrary $c \in \mathbf{Q}^+$, but we suggest an alternate proof for negative c in exercise 6.97.

6.87 Example (Numerical calculation of e) It follows from the last example that

$$\lim\left\{\left(1+\frac{1}{n}\right)^n\right\} = e.$$

I wrote a Maple procedure to calculate e by using this fact. The procedure `limcalc(n)` below calculates

$$\left(1+\frac{1}{100^n}\right)^{100^n},$$

and I have printed out the results for n = 1,2,...,6.

```
> limcalc := n -> (1+ .01^n)^(100^n);
```

$$limcalc := n \to (1 + .01^n)^{(100^n)}$$

```
> limcalc(1);
```

$$2.704813829$$

```
> limcalc(2);
```

$$2.718145927$$

```
> limcalc(3);
```

$$2.718280469$$

```
> limcalc(4);
```

$$2.718281815$$

```
> limcalc(5);
```

$$1.$$

```
> limcalc(6);
```

$$1.$$

6.88 Exercise. From my computer calculations it appears that

$$\lim\left\{\left(1+\frac{1}{n}\right)^n\right\} = 1.$$

Explain what has gone wrong. What can I conclude about the value of e from my program?

6.89 Example. Actually, Maple is smart enough to find the limit, and does so with the commands below. The command `evalf` returns the decimal approximation of its argument.

```
> limit( (1+1/n)^n,n=infinity);
```

e

```
> evalf(%);
```

2.718281828

6.90 Entertainment ($\lim\{n^{\frac{1}{n}}\}$.) Find the limit of the sequence $\{n^{\frac{1}{n}}\}$, or else show that the sequence diverges.

6.91 Example (Compound interest.) The previous exercise has the following interpretation.

Suppose that A dollars is invested at r% annual interest, compounded n times a year. The value of the investment at any time t is calculated as follows:

Let $T = (1/n)$ year, and let A_n^k be the value of the investment at time kT. Then

$$\begin{aligned} A_n^0 &= A \\ A_n^1 &= A_n^0 + \frac{r}{100n}A_n^0 = (1+\frac{r}{100n})A \\ A_n^2 &= A_n^1 + \frac{r}{100n}A_n^1 = (1+\frac{r}{100n})^2 A \end{aligned} \tag{6.92}$$

and in general

$$A_n^k = A_n^{k-1} + \frac{r}{100n}A_n^{k-1} = (1+\frac{r}{100n})^k A. \tag{6.93}$$

The value of the investment does not change during the time interval $kT < t < (k+1)T$. For example, if V_n denotes the value of one dollar invested for one year at $r\%$ annual rate of interest with the interest compounded n times a year, then

$$V_n = A_n^n = \left(1 + \frac{r}{100n}\right)^n.$$

Thus it follows from our calculation that if one dollar is invested for one year at $r\%$ annual rate of interest, with the interest compounded "infinitely often" or "continuously", then the value of the investment at the end of the year will be

$$\lim\left\{\left(1 + \frac{r}{100n}\right)^n\right\} = e^{\frac{r}{100}} \text{ dollars.}$$

If the rate of interest is 100%, then the value of the investment is e dollars, and the investor should expect to get \$2.71 from the bank.

This example was considered by Jacob Bernoulli in 1685. Bernoulli was able to show that $\lim\left\{\left(1 + \frac{1}{n}\right)^n\right\} < 3$.[8, pp94-97]

6.94 Exercise. Calculate the following limits.

a) $\lim\{(1 + \frac{3}{n})^{2n}\}$.

b) $\lim\{(1 + \frac{1}{3n})^{2n}\}$.

6.95 Exercise.

a) Use the formula for a finite geometric series,

$$1+(1-a)+(1-a)^2+\cdots+(1-a)^{n-1}=\frac{1-(1-a)^n}{1-(1-a)}$$

to show that

$$(1-a)^n \geq 1-na \text{ whenever } 0<a<1. \tag{6.96}$$

b) Let $c \in \mathbf{R}^+$ Use inequality (6.96) to show that

$$\left(1-\frac{c}{n^2}\right)^n \geq 1-\frac{c}{n}$$

for all $n \in \mathbf{Z}^+$ such that $n > \sqrt{c}$.

c) Prove that $\lim\{(1-\frac{c}{n^2})^n\} = 1$ for all $c \in \mathbf{R}^+$.

6.97 Exercise. Let $c \in \mathbf{Q}^+$. Use exercise 6.95 to show that

$$\lim\left\{\left(1-\frac{c}{n}\right)^n\right\} = e^{-c}.$$

(Hence we have $\lim\{(1+\frac{c}{n})^n\} = e^c$ for all $c \in \mathbf{Q}$.)
Hint: Note that $(1-z) = (\frac{1-z^2}{1+z})$ for all real numbers $z \neq -1$.

Chapter 7

Still More Area Calculations

7.1 Area Under a Monotonic Function

7.1 Theorem. *Let f be a monotonic function from the interval $[a, b]$ to $\mathbf{R}_{\geq 0}$. Let $\{P_n\}$ be a sequence of partitions of $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and let*

$$A_a^b f = \alpha\{(x, y) \in \mathbf{R}^2 : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}$$

Then

$$\{\alpha\Big(I_a^b(f, P_n)\Big)\} \to A_a^b f$$

and

$$\{\alpha\Big(O_a^b(f, P_n)\Big)\} \to A_a^b f.$$

(The notation here is the same as in theorem 5.40 and exercise 5.47.)

Proof: We noted in theorem 5.40 and exercise 5.47 that

$$0 \leq \alpha\Big(O_a^b(f, P_n)\Big) - \alpha\Big(I_a^b(f, P_n)\Big) \leq \mu(P_n) \cdot |f(b) - f(a)|. \tag{7.2}$$

Since

$$\begin{aligned} \lim \{\mu(P_n) \cdot |f(b) - f(a)|\} &= |f(b) - f(a)| \lim\{\mu(P_n)\} \\ &= |f(b) - f(a)| \cdot 0 = 0, \end{aligned}$$

we conclude from the squeezing rule that

$$\lim \Big\{\alpha\Big(O_a^b(f, P_n)\Big) - \alpha\Big(I_a^b(f, P_n)\Big)\Big\} = 0. \tag{7.3}$$

We also have by (5.43) that

$$\alpha\Big(I_a^b(f,P_n)\Big) \leq A_a^b f \leq \alpha\Big(O_a^b(f,P_n)\Big),$$

so that

$$0 \leq A_a^b f - \alpha\Big(I_a^b(f,P_n)\Big) \leq \alpha\Big(O_a^b(f,P_n)\Big) - \alpha\Big(I_a^b(f,P_n)\Big).$$

By (7.3) and the squeezing rule

$$\lim\Big\{A_a^b f - \alpha\Big(I_a^b(f,P_n)\Big)\Big\} = 0,$$

and hence

$$\lim\Big\{\alpha\Big(I_a^b(f,P_n)\Big)\Big\} = A_a^b f.$$

Also,

$$\begin{aligned}\lim\Big\{\alpha\Big(O_a^b(f,P_n)\Big)\Big\} &= \lim\Big\{\alpha\Big(I_a^b(f,P_n)\Big) + \Big(\alpha(O_a^b(f,P_n)) - \alpha(I_a^b(f,P_n))\Big)\Big\}\\ &= \lim\Big\{\alpha\Big(I_a^b(f,P_n)\Big)\Big\}\\ &\qquad + \lim\Big\{\Big(\alpha(O_a^b(f,P_n)) - \alpha(I_a^b(f,P_n))\Big)\Big\}\\ &= A_a^b f + 0 = A_a^b f. \ \|\end{aligned}$$

7.4 Definition (Riemann sum). Let $P = \{x_0, x_1, \cdots, x_n\}$ be a partition for an interval $[a,b]$. A *sample* for P is a finite sequence $S = \{s_1, s_2, \cdots, s_n\}$ of numbers such that $s_i \in [x_{i-1}, x_i]$ for $1 \leq i \leq n$. If f is a function from $[a,b]$ to $\mathbf{R}$, and P is a partition for $[a,b]$ and S is a sample for P, we define

$$\sum(f,P,S) = \sum_{i=1}^{n} f(s_i)(x_i - x_{i-1})$$

and we call $\sum(f,P,S)$ a *Riemann sum* for f, P and S. We will sometimes write $\sum([f(t)],P,S)$ instead of $\sum(f,P,S)$.

7.5 Example. If f is an increasing function from $[a,b]$ to $\mathbf{R}_{\geq 0}$, and $P = \{x_0, \cdots, x_n\}$ is a partition of $[a,b]$, and $S_l = \{x_0, \cdots, x_{n-1}\}$, then

$$\sum(f,P,S_l) = \alpha(I_a^b(f,P)).$$

If $S_r = \{x_1, x_2, \cdots, x_n\}$, then

$$\sum(f,P,S_r) = \alpha(O_a^b(f,P)).$$

If $S_m = \{\frac{x_0 + x_1}{2}, \cdots, \frac{x_{n-1} + x_n}{2}\}$ then

$$\sum(f, P, S_m) = \sum_{i=1}^{n} f\Big(\frac{x_{i-1} + x_i}{2}\Big)(x_i - x_{i-1})$$

is some number between $\alpha(I_a^b(f, P))$ and $\alpha(O_a^b(f, P))$.

7.6 Theorem (Area theorem for monotonic functions.) *Let f be a monotonic function from the interval $[a, b]$ to $\mathbf{R}_{\geq 0}$. Then for every sequence $\{P_n\}$ of partitions of $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and for every sequence $\{S_n\}$ where S_n is a sample for P_n, we have*

$$\{\sum(f, P_n, S_n)\} \to A_a^b f.$$

Proof: We will consider the case where f is increasing. The case where f is decreasing is similar.

For each partition $P_n = \{x_0, \cdots, x_m\}$ and sample $S_n = \{s_1, \cdots, s_m\}$, we have for $1 \leq i \leq m$

$$\begin{aligned} x_{i-1} \leq s_i \leq x_i &\implies f(x_{i-1}) \leq f(s_i) \leq f(x_i) \\ &\implies f(x_{i-1})(x_i - x_{i-1}) \leq f(s_i)(x_i - x_{i-1}) \leq f(x_i)(x_i - x_{i-1}). \end{aligned}$$

Hence

$$\sum_{i=1}^{m} f(x_{i-1})(x_i - x_{i-1}) \leq \sum_{i=1}^{m} f(s_i)(x_i - x_{i-1}) \leq \sum_{i=1}^{m} f(x_i)(x_i - x_{i-1}),$$

i.e.,

$$\alpha\Big(I_a^b(f, P_n)\Big) \leq \sum(f, P_n, S_n) \leq \alpha\Big(O_a^b(f, P_n)\Big).$$

By theorem 7.1 we have

$$\{\alpha\Big(I_a^b(f, P_n)\Big)\} \to A_a^b f,$$

and

$$\{\alpha\Big(O_a^b(f, P_n)\Big)\} \to A_a^b f,$$

so by the squeezing rule,

$$\{\sum(f, P_n, S_n)\} \to A_a^b f.$$

7.2 Calculation of Area under Power Functions

7.7 Lemma. *Let r be a rational number such that $r \neq -1$. Let a be a real number with $a > 1$. Then*

$$A_1^a[t^r] = (a^{r+1} - 1) \lim \left\{ \frac{a^{\frac{1}{n}} - 1}{a^{\frac{r+1}{n}} - 1} \right\}.$$

(For the purposes of this lemma, we will assume that the limit exists. In theorem 7.10 we will prove that the limit exists.)

Proof: Let n be a generic element of $\mathbf{Z}^+$. To simplify the notation, I will write

$$p = a^{\frac{1}{n}}, \text{ (so } p > 1).$$

Let

$$P_n = \{1, a^{\frac{1}{n}}, a^{\frac{2}{n}}, \cdots, a^{\frac{n}{n}}\} = \{1, p, p^2, \cdots, p^n\} = \{x_0, x_1, x_2, \cdots, x_n\}$$

and let

$$S_n = \{1, p, p^2, \cdots, p^{n-1}\} = \{s_1, s_2, s_3 \cdots, s_n\}.$$

Then for $1 \leq i \leq n$

$$x_i - x_{i-1} = p^i - p^{i-1} = p^{i-1}(p-1),$$

so

$$\mu(P_n) = p^{n-1}(p-1) \leq p^n(p-1) = a\left(a^{\frac{1}{n}} - 1\right).$$

It follows by the nth root rule (theorem 6.48) that $\{\mu(P_n)\} \to 0$. Hence it follows from theorem 7.6 that

$$A_1^a[t^r] = \lim\left(\sum([t^r], P_n, S_n)\right). \tag{7.8}$$

Now

$$\begin{aligned}\sum([t^r], P_n, S_n) &= \sum_{i=1}^{n} s_i^r (x_i - x_{i-1}) \\ &= \sum_{i=1}^{n} (p^{(i-1)})^r p^{i-1}(p-1)\end{aligned}$$

$$
\begin{aligned}
&= (p-1)\sum_{i=1}^{n}\left(p^{r+1}\right)^{(i-1)} \qquad (7.9)\\
&= (p-1)\left(\frac{p^{(r+1)n}-1}{p^{r+1}-1}\right) = \left((p^n)^{r+1}-1\right)\left(\frac{p-1}{p^{r+1}-1}\right)\\
&= \left(a^{r+1}-1\right)\frac{(a^{\frac{1}{n}}-1)}{(a^{\frac{r+1}{n}}-1)}.
\end{aligned}
$$

Here we have used the formula for a finite geometric series. Thus, from (7.8)

$$
\begin{aligned}
A_1^a[t^r] &= \lim\left\{(a^{r+1}-1)\frac{(a^{\frac{1}{n}}-1)}{(a^{\frac{r+1}{n}}-1)}\right\}\\
&= (a^{r+1}-1)\lim\left\{\frac{a^{\frac{1}{n}}-1}{a^{\frac{r+1}{n}}-1}\right\}. \parallel
\end{aligned}
$$

Now we want to calculate the limit appearing in the previous lemma. In order to do this it will be convenient to prove a few general limit theorems.

7.10 Theorem. *Let $\{x_n\}$ be a sequence of positive numbers such that $\{x_n\} \to 1$ and $x_n \neq 1$ for all $n \in \mathbf{Z}^+$. Let β be any rational number. Then*

$$\left\{\frac{x_n^\beta - 1}{x_n - 1}\right\} \to \beta.$$

Proof: Suppose $x_n \neq 1$ for all n, and $\{x_n\} \to 1$.
Case 1: Suppose $\beta = 0$. Then the conclusion clearly follows.
Case 2: Suppose $\beta \in \mathbf{Z}^+$. Then by the formula for a geometric series

$$\frac{x_n^\beta - 1}{x_n - 1} = 1 + x_n + \cdots + x_n^{\beta-1}.$$

By the sum theorem and many applications of the product theorem we conclude that

$$
\begin{aligned}
\lim\left\{\frac{x_n^\beta - 1}{x_n - 1}\right\} &= \lim\{1\} + \lim\{x_n\} + \cdots + \lim\{x_n^{\beta-1}\}\\
&= 1 + 1 + 1 + \cdots + 1\\
&= \beta.
\end{aligned}
$$

Case 3: Suppose $\beta \in \mathbf{Z}^-$. Let $\gamma = -\beta$. Then $\gamma \in \mathbf{Z}^+$, so by Case 2 we get

$$\begin{aligned}
\lim\Big\{\frac{x_n^\beta - 1}{x_n - 1}\Big\} &= \lim\Big\{\frac{x_n^\gamma(x_n^\beta - 1)}{x_n^\gamma(x_n - 1)}\Big\} = \lim\Big\{\frac{1 - x_n^\gamma}{x_n^\gamma(x_n - 1)}\Big\} \\
&= \lim\Big\{\frac{1}{-x_n^\gamma}\Big(\frac{x_n^\gamma - 1}{x_n - 1}\Big)\Big\} \\
&= \lim\Big\{\frac{1}{-x_n^\gamma}\Big\}\lim\Big\{\frac{x_n^\gamma - 1}{x_n - 1}\Big\} \\
&= \frac{1}{-1}\cdot\gamma = -\gamma = \beta.
\end{aligned}$$

Case 4: Suppose $\beta = \dfrac{p}{q}$ where $q \in \mathbf{Z}^+$ and $p \in \mathbf{Z}$. Let $y_n = x_n^{\frac{1}{q}}$. Then

$$\frac{x_n^\beta - 1}{x_n - 1} = \frac{x_n^{\frac{p}{q}} - 1}{x_n - 1} = \frac{y_n^p - 1}{y_n^q - 1} = \frac{\Big(\dfrac{y_n^p - 1}{y_n - 1}\Big)}{\Big(\dfrac{y_n^q - 1}{y_n - 1}\Big)}.$$

Now if we could show that $\{y_n\} \to 1$, it would follow from this formula that

$$\lim\Big\{\frac{x_n^\beta - 1}{x_n - 1}\Big\} = \frac{\lim\Big\{\frac{y_n^p - 1}{y_n - 1}\Big\}}{\lim\Big\{\frac{y_n^q - 1}{y_n - 1}\Big\}} = \frac{p}{q} = \beta.$$

The next lemma shows that $\{y_n\} \to 1$ and completes the proof of theorem 7.10.

7.11 Lemma. *Let $\{x_n\}$ be a sequence of positive numbers such that $\{x_n\} \to 1$, and $\{x_n\} \neq 1$ for all $n \in \mathbf{Z}^+$. Then for each q in $\mathbf{Z}^+$, $\{x_n^{\frac{1}{q}}\} \to 1$.*

Proof: Let $\{x_n\}$ be a sequence of positive numbers such that $\{x_n\} \to 1$. Let $y_n = x_n^{\frac{1}{q}}$ for each n in $\mathbf{Z}^+$. We want to show that $\{y_n\} \to 1$. By the formula for a finite geometric series

$$1 + y_n + \cdots + y_n^{q-1} = \frac{(1 - y_n^q)}{1 - y_n} = \frac{(1 - x_n)}{1 - y_n}$$

so

$$(1 - y_n) = \frac{(1 - x_n)}{1 + y_n + \cdots + y_n^{q-1}}.$$

Now

$$0 \leq |1 - y_n| = \frac{|1 - x_n|}{|1 + y_n + \cdots + y_n^{q-1}|} = \frac{|1 - x_n|}{1 + y_n + \cdots + y_n^{q-1}} \leq |1 - x_n|.$$

Since $\{x_n\} \to 1$, we have $\lim\{|1-x_n|\} = 0$, so by the squeezing rule $\lim\{|1-y_n|\} = 0$, and hence

$$\lim\{y_n\} = 1. \; \|$$

7.12 Lemma (Calculation of $A_1^b[t^r]$.) *Let b be a real number with $b > 1$, and let $r \in \mathbf{Q} \setminus \{-1\}$. Then*

$$A_1^b[t^r] = \frac{b^{r+1} - 1}{r + 1}.$$

Proof: By lemma 7.7,

$$A_1^b[t^r] = (b^{r+1} - 1) \lim \left\{ \frac{b^{\frac{1}{n}} - 1}{b^{\frac{r+1}{n}} - 1} \right\}.$$

By theorem 7.10,

$$\lim \left\{ \frac{b^{\frac{1}{n}} - 1}{b^{\frac{r+1}{n}} - 1} \right\} = \lim \left\{ \frac{1}{\frac{b^{\frac{r+1}{n}} - 1}{b^{\frac{1}{n}} - 1}} \right\} = \frac{\lim\{1\}}{\lim \left\{ \frac{b^{\frac{r+1}{n}} - 1}{b^{\frac{1}{n}} - 1} \right\}} = \frac{1}{r + 1},$$

and putting these results together, we get

$$A_1^b[t^r] = \frac{b^{r+1} - 1}{r + 1}. \; \|$$

7.13 Lemma. *Let $r \in \mathbf{Q}$, and let $a, c \in \mathbf{R}^+$, with $1 < c$. Then*

$$A_a^{ca}[t^r] = a^{r+1} A_1^c[t^r].$$

Proof: If

$$P = \{x_0, x_1, \ldots, x_n\}$$

is a partition of $[1, c]$, let

$$aP = \{ax_0, ax_1, \ldots, ax_n\}$$

be the partition of $[a, ca]$ obtained by multiplying the points of P by a. Then

$$\mu(aP) = a\mu(P). \tag{7.14}$$

If

$$S = \{s_1, s_2, \ldots, s_n\}$$

is a sample for P, let

$$aS = \{as_1, as_2, \ldots, as_n\}$$

be the corresponding sample for aP. Then

$$\begin{aligned}
\sum([t^r], aP, aS) &= \sum_{i=1}^{n} (as_i)^r (ax_i - ax_{i-1}) \\
&= \sum_{i=1}^{n} a^r s_i^r a(x_i - x_{i-1}) \\
&= a^{r+1} \sum_{i=1}^{n} s_i^r (x_i - x_{i-1}) \\
&= a^{r+1} \sum([t^r], P, S).
\end{aligned}$$

Let $\{P_n\}$ be a sequence of partitions of $[1, c]$ such that $\{\mu(P_n)\} \to 0$, and for each $n \in \mathbf{Z}^+$ let S_n be a sample for P_n. It follows from (7.14) that $\{\mu(aP_n)\} \to 0$. By the area theorem for monotonic functions (theorem 7.6), we have

$$\left\{\sum([t^r], P_n, S_n)\right\} \to A_1^c[t^r] \text{ and } \left\{\sum([t^r], aP_n, aS_n)\right\} \to A_a^{ca}[t^r].$$

Thus

$$\begin{aligned}
A_a^{ca}[t^r] &= \lim\{\sum([t^r], aP_n, aS_n)\} \\
&= \lim\{a^{r+1} \sum([t^r], P_n, S_n)\} = a^{r+1} \lim\{\sum([t^r], P_n, S_n)\} \\
&= a^{r+1} A_1^c[t^r]. \ \|
\end{aligned}$$

7.15 Theorem (Calculation of $A_a^b[t^r]$.) *Let $a, b \in \mathbf{R}^+$ with $a < b$, and let $r \in \mathbf{Q}$. Then*

$$A_a^b[t^r] = \begin{cases} \frac{b^{r+1} - a^{r+1}}{r+1} & \text{if } r \neq -1 \\ \ln(b) - \ln(a) & \text{if } r = -1. \end{cases}$$

Proof: The result for the case $r = -1$ was proved in theorem 5.76. The case $r \neq -1$ is done in the following exercise.

7.16 Exercise. Use the two previous lemmas to prove theorem 7.15 for the case $r \neq -1$.

Remark: In the proof of lemma 7.7, we did not use the assumption $r \neq -1$ until line (7.9). For $r = -1$ equation (7.9) becomes

$$\sum([t^{-1}], P_n, S_n) = n(a^{\frac{1}{n}} - 1).$$

Since in this case $\{\sum([t^{-1}], P_n, S_n)\} \to A_1^a[\frac{1}{t}] = \ln(a)$,we conclude that

$$\lim\{n(a^{\frac{1}{n}} - 1)\} = \ln(a) \text{ for all } a > 1. \tag{7.17}$$

This formula give us method of calculating logarithms by taking square roots. We know $2^n(a^{\frac{1}{2^n}} - 1)$ will be near to $\ln(a)$ when n is large, and $a^{\frac{1}{2^n}}$ can be calculated by taking n successive square roots. On my calculator, I pressed the following sequence of keys

$$2 \underbrace{\sqrt{}\sqrt{}\cdots\sqrt{}}_{15 \text{ times}} -1 \; = \; \underbrace{\times 2 \quad \times 2 \quad \cdots \quad \times 2}_{15 \text{ times}} \; =$$

and got the result 0.693154611. My calculator also says that
$\ln(2) = 0.69314718$. It appears that if I know how to calculate square roots, then I can calculate logarithms fairly easily.

7.18 Exercise. Let r be a non-negative rational number, and let $b \in \mathbf{R}^+$. Show that

$$A_0^b[t^r] = \frac{b^{r+1}}{r+1}.$$

Where in your proof do you use the fact that $r \geq 0$?

Chapter 8

Integrable Functions

8.1 Definition of the Integral

If f is a monotonic function from an interval $[a, b]$ to $\mathbf{R}_{\geq 0}$, then we have shown that for every sequence $\{P_n\}$ of partitions on $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and every sequence $\{S_n\}$ such that for all $n \in \mathbf{Z}^+$ S_n is a sample for P_n, we have

$$\{\sum(f, P_n, S_n)\} \to A_a^b f.$$

8.1 Definition (Integral.) Let f be a bounded function from an interval $[a, b]$ to $\mathbf{R}$. We say that f is *integrable on* $[a, b]$ if there is a number V such that for every sequence of partitions $\{P_n\}$ on $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and every sequence $\{S_n\}$ where S_n is a sample for P_n

$$\{\sum(f, P_n, S_n)\} \to V.$$

If f is integrable on $[a, b]$ then the number V just described is denoted by $\displaystyle\int_a^b f$ and is called "the integral from a to b of f." Notice that by our definition an integrable function is necessarily bounded.

The definition just given is essentially due to Bernhard Riemann(1826–1866), and first appeared around 1860[39, pages 239 ff]. The symbol $\displaystyle\int$ was introduced by Leibniz sometime around 1675[15, vol 2, p242]. The symbol is a form of the letter s, standing for *sum* (in Latin as well as in English.) The practice of attaching the limits a and b to the integral sign was introduced by

Joseph Fourier around 1820. Before this time the limits were usually indicated by words.

We can now restate theorems 7.6 and 7.15 as follows:

8.2 Theorem (Monotonic functions are integrable I.) *If f is a monotonic function on an interval $[a, b]$ with non-negative values, then f is integrable on $[a, b]$ and*

$$\int_a^b f = A_a^b f = \alpha(\{(x, y) : a \le x \le b \text{ and } 0 \le y \le f(x)\}).$$

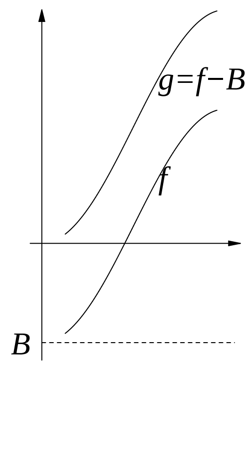
8.3 Theorem (Integrals of power functions.) *Let $r \in \mathbf{Q}$, and let a, b be real numbers such that $0 < a \le b$. Let $f_r(x) = x^r$ for $a \le x \le b$. Then*

$$\int_a^b f_r = \begin{cases} \dfrac{b^{r+1} - a^{r+1}}{r+1} & \text{if } r \in \mathbf{Q} \setminus \{-1\} \\ \ln(b) - \ln(a) & \text{if } r = -1. \end{cases}$$

In general integrable functions may take negative as well as positive values and in these cases $\int_a^b f$ does not represent an area.

The next theorem shows that monotonic functions are integrable even if they take on negative values.

8.4 Example (Monotonic functions are integrable II.) Let f be a monotonic function from an interval $[a, b]$ to $\mathbf{R}$. Let B be a non-positive number such that $f(x) \ge B$ for all $x \in [a, b]$. Let $g(x) = f(x) - B$.



Then g is a monotonic function from $[a, b]$ to $\mathbf{R}_{\geq 0}$. Hence by theorem 7.6, g is integrable on $[a, b]$ and $\int_a^b g = A_a^b(g)$. Now let $\{P_n\}$ be a sequence of partitions of $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and let $\{S_n\}$ be a sequence such that for each n in $\mathbf{Z}^+$, S_n is a sample for P_n. Then

$$\{\sum(g, P_n, S_n)\} \to A_a^b(g). \tag{8.5}$$

If $P_n = \{x_0, \cdots, x_m\}$ and $S_n = \{s_1, \cdots, s_m\}$ then

$$\begin{aligned}
\sum(g, P_n, S_n) &= \sum_{i=1}^{m} g(s_i)(x_i - x_{i-1}) \\
&= \sum_{i=1}^{m} \Big(f(s_i) - B\Big)(x_i - x_{i-1}) \\
&= \sum_{i=1}^{m} f(s_i)(x_i - x_{i-1}) - B\sum_{i=1}^{m}(x_i - x_{i-1}) \\
&= \sum(f, P_n, S_n) - B(b - a).
\end{aligned}$$

Thus by (8.5)

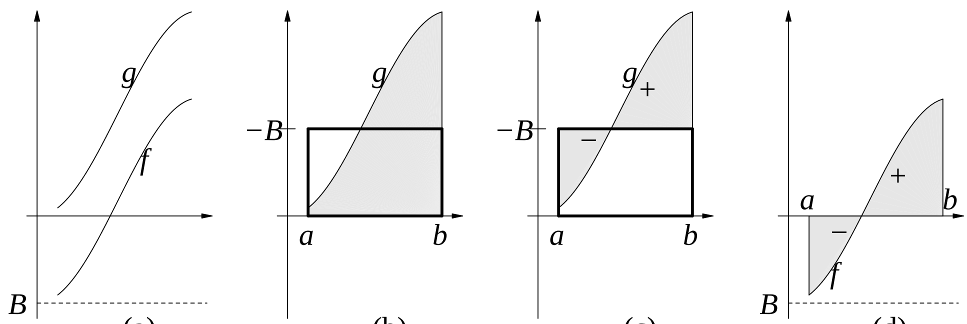
$$\{\sum(f, P_n, S_n) - B(b - a)\} \to A_a^b(g).$$

If we use the fact that $\{B(b - a)\} \to B(b - a)$, and then use the sum theorem for limits of sequences, we get

$$\{\sum(f, P_n, S_n)\} \to A_a^b(g) + B(b - a).$$

It follows from the definition of integrable functions that f is integrable on $[a, b]$ and

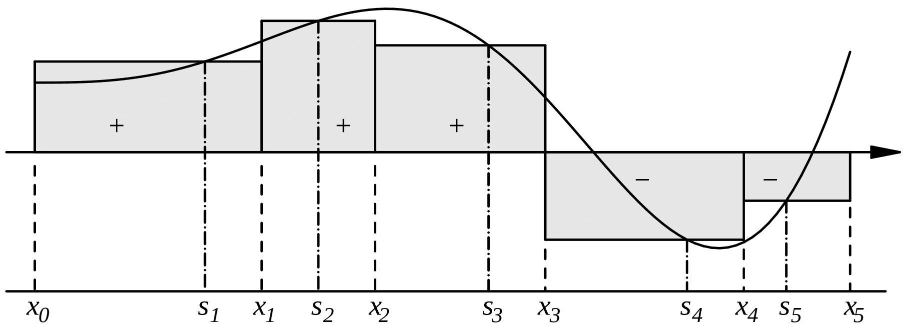
$$\int_a^b f = A_a^b(g) + B(b - a) = \int_a^b g + B(b - a) = \int_a^b g - |B|(b - a).$$



(a) (b) (c) (d)

Thus in figure b, $\int_a^b f$ represents the shaded area with the area of the thick box subtracted from it, which is the same as the area of the region marked "+" in figures c and d, with the area of the region marked "−" subtracted from it.

The figure represents a geometric interpretation for a Riemann sum. In the figure



$$f(s_i) > 0 \text{ for } i = 1, 2, 3, \quad f(s_i) < 0 \text{ for } i = 4, 5.$$

$$\sum_{i=1}^{3} f(s_i)(x_i - x_{i-1})$$

is the area of $\bigcup_{i=1}^{3} B\Big(x_{i-1}, x_i\colon 0, f(s_i)\Big)$ and

$$\sum_{i=4}^{5} f(s_i)(x_i - x_{i-1})$$

is the *negative* of the area of

$$\bigcup_{i=4}^{5} B(x_{i-1}, x_i : f(s_i), 0).$$

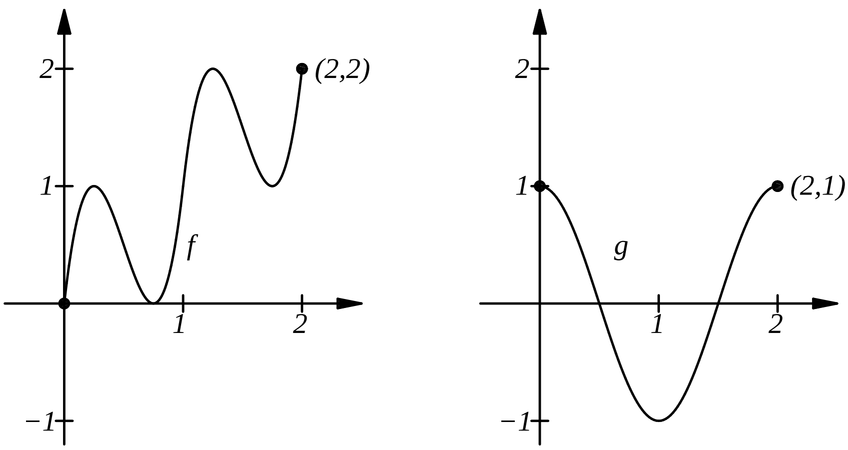
In general you should think of $\int_a^b f$ as the difference $\alpha(S^+) - \alpha(S^-)$ where

$$S^+ = \{(x, y)\colon a \le x \le b \text{ and } 0 \le y \le f(x)\}$$

and

$$S^- = \{(x, y): a \le x \le b \text{ and } f(x) \le y \le 0\}.$$

8.6 Exercise. The graphs of two functions f, g from $[0, 2]$ to $\mathbf{R}$ are sketched below.



Let

$$F(x) = \big(f(x)\big)^2 \text{ for } 0 \le x \le 2, \quad G(x) = \big(g(x)\big)^2 \text{ for } 0 \le x \le 2.$$

Which is larger:

a) $\displaystyle\int_0^1 f$ or $\displaystyle\int_0^1 F$?

b) $\displaystyle\int_0^1 g$ or $\displaystyle\int_0^1 G$?

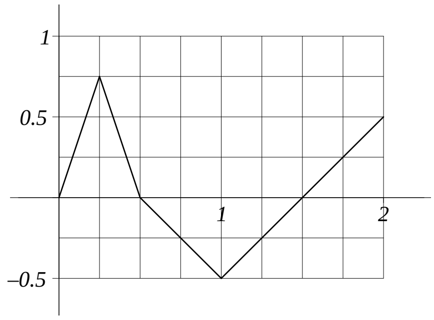
c) $\displaystyle\int_0^1 f$ or $\displaystyle\int_0^1 g$?

d) $\displaystyle\int_0^{1/2} g$ or $\displaystyle\int_0^{1/2} G$?

e) $\displaystyle\int_0^2 g$ or $\displaystyle\int_0^2 G$?

Explain how you decided on your answers. Your explanations may be informal, but they should be convincing.

8.7 Exercise. Below is the graph of a function g. By looking at the graph of g estimate the following integrals. (No explanation is necessary.)



Graph of g

a) $\displaystyle\int_{\frac{1}{4}}^{\frac{3}{4}} g.$

b) $\displaystyle\int_{1}^{2} g.$

c) $\displaystyle\int_{0}^{\frac{3}{4}} g.$

8.8 Exercise. Sketch the graph of one function f satisfying all four of the following conditions.

a) $\displaystyle\int_{0}^{1} f = 1.$

b) $\displaystyle\int_{0}^{2} f = -1.$

c) $\displaystyle\int_{0}^{3} f = 0.$

d) $\displaystyle\int_{0}^{4} f = 1.$

8.2 Properties of the Integral

8.9 Definition (Operations on functions.) Let $f: S \to \mathbf{R}$ and $g: T \to \mathbf{R}$ be functions where S, T are sets. Let $c \in \mathbf{R}$. We define functions $f \pm g$, fg, cf, $\frac{f}{g}$ and $|f|$ as follows:

$(f+g)(x) = f(x) + g(x)$ for all $x \in S \cap T$.
$(f-g)(x) = f(x) - g(x)$ for all $x \in S \cap T$.
$(fg)(x) = f(x)g(x)$ for all $x \in S \cap T$.
$(cf)(x) = c \cdot f(x)$ for all $x \in S$.
$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}$ for all $x \in S \cap T$ such that $g(x) \neq 0$.
$|f|(x) = |f(x)|$ for all $x \in S$.

Remark: These operations of addition, subtraction, multiplication and division for functions satisfy the associative, commutative and distributive laws that you expect them to. The proofs are straightforward and will be omitted.

8.10 Definition (Partition-sample sequence.) Let $[a, b]$ be an interval. By a *partition-sample sequence* for $[a, b]$ I will mean a pair of sequences $(\{P_n\}, \{S_n\})$ where $\{P_n\}$ is a sequence of partitions of $[a, b]$ such that $\{\mu(P_n)\} \to 0$, and for each n in $\mathbf{Z}^+$, S_n is a sample for P_n.

8.11 Theorem (Sum theorem for integrable functions.) *Let f, g be integrable functions on an interval $[a, b]$. Then $f \pm g$ and cf are integrable on $[a, b]$ and*

$$\int_a^b (f \pm g) = \int_a^b f \pm \int_a^b g,$$

and

$$\int_a^b cf = c \int_a^b f.$$

Proof: Suppose f and g are integrable on $[a, b]$. Let $(\{P_n\}, \{S_n\})$ be a partition-sample sequence for $[a, b]$. If $P_n = \{x_0, \cdots, x_m\}$ and $S_n = \{s_1, \cdots, s_m\}$, then

$$\sum(f \pm g, P_n, S_n) = \sum_{i=1}^{m} (f \pm g)(s_i)(x_i - x_{i-1})$$

$$\begin{aligned}
&= \sum_{i=1}^{m} \Big(f(s_i) \pm g(s_i)\Big)(x_i - x_{i-1}) \\
&= \sum_{i=1}^{m} f(s_i)(x_i - x_{i-1}) \pm \sum_{i=1}^{m} g(s_i)(x_i - x_{i-1}) \\
&= \sum(f, P_n, S_n) \pm \sum(g, P_n, S_n).
\end{aligned}$$

Since f and g are integrable, we have

$$\{\sum(f, P_n, S_n)\} \to \int_a^b f \text{ and } \{\sum(g, P_n, S_n)\} \to \int_a^b g.$$

By the sum theorem for sequences,

$$\{\sum(f \pm g), P_n, S_n)\} = \{\sum(f, P_n, S_n) \pm \sum(g, P_n, S_n)\} \to \int_a^b f \pm \int_a^b g.$$

Hence $f \pm g$ is integrable and $\int_a^b (f \pm g) = \int_a^b f \pm \int_a^b g$. The proof of the second statement is left as an exercise.

8.12 Notation ($\int_a^b f(t)\, dt$) If f is integrable on an interval $[a, b]$ we will sometimes write $\int_a^b f(x)\, dx$ instead of $\int_a^b f$. The "x" in this expression is a dummy variable, but the "d" is a part of the notation and may not be replaced by another symbol. This notation will be used mainly in cases where no particular name is available for f. Thus

$$\int_1^2 t^3 + 3t\, dt \text{ or } \int_1^2 x^3 + 3x\, dx \text{ or } \int_1^2 (x^3 + 3x)dx$$

means $\int_1^2 F$ where F is the function on $[1, 2]$ defined by $F(t) = t^3 + 3t$ for all $t \in [1, 2]$. The "d" here stands for difference, and dx is a ghost of the differences $x_i - x_{i-1}$ that appear in the approximations for the integral. The dx notation is due to Leibniz.

8.13 Example. Let

$$f(x) = (x-1)^2 - \frac{1}{x} + \frac{3}{\sqrt{x}} = x^2 - 2x + x^0 - \frac{1}{x} + 3x^{-\frac{1}{2}}.$$

This function is integrable over every closed bounded subinterval of $(0,\infty)$, since it is a sum of five functions that are known to be integrable. By several applications of the sum theorem for integrals we get

$$\begin{aligned}\int_1^2 f &= \int_1^2 (x^2-2x+1-\frac{1}{x}+3x^{-\frac{1}{2}})dx \\ &= \left(\frac{2^3-1^3}{3}\right)-2\left(\frac{2^2-1^2}{2}\right)+\left(\frac{2^1-1^1}{1}\right)-\ln(2)+3\left(\frac{2^{\frac{1}{2}}-1^{\frac{1}{2}}}{\frac{1}{2}}\right) \\ &= \frac{7}{3}-3+1-\ln(2)+6(\sqrt{2}-1)=-\frac{17}{3}-\ln(2)+6\sqrt{2}.\end{aligned}$$

8.14 Exercise. Calculate the following integrals.

a) $\displaystyle\int_1^a (2-x)^2dx$. Here $a>1$.

b) $\displaystyle\int_1^4 \sqrt{x}-\frac{1}{x^2}\,dx$.

c) $\displaystyle\int_1^{27} x^{-\frac{1}{3}}dx$.

d) $\displaystyle\int_0^{27} x^{-\frac{1}{3}}dx$.

e) $\displaystyle\int_1^2 \frac{x+1}{x}dx$.

f) $\displaystyle\int_a^b M\,dx$. Here $a\le b$, and M denotes a constant function.

8.15 Theorem (Inequality theorem for integrals.) *Let f and g be integrable functions on the interval $[a,b]$ such that*

$$f(x)\le g(x) \text{ for all } x\in[a,b].$$

Then

$$\int_a^b f \le \int_a^b g.$$

8.16 Exercise. Prove the inequality theorem for integrals.

8.17 Corollary. *Let f be an integrable function on the interval $[a,b]$. Suppose $|f(x)| \le M$ for all $x \in [a,b]$. Then*

$$\left| \int_a^b f \right| \le M(b-a).$$

Proof: We have

$$-M \le f(x) \le M \text{ for all } x \in [a,b].$$

Hence by the inequality theorem for integrals

$$\int_a^b -M \le \int_a^b f \le \int_a^b M.$$

Hence

$$-M(b-a) \le \int_a^b f \le M(b-a).$$

It follows that

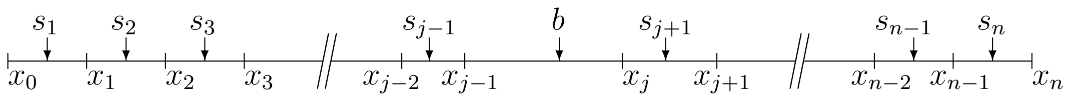
$$\left| \int_a^b f \right| \le M(b-a). \;\|$$

8.18 Theorem. *Let a, b, c be real numbers with $a < b < c$, and let f be a function from $[a,c]$ to $\mathbf{R}$. Suppose f is integrable on $[a,b]$ and f is integrable on $[b,c]$. Then f is integrable on $[a,c]$ and $\int_a^c f = \int_a^b f + \int_b^c f$.*

Proof: Since f is integrable on $[a,b]$ and on $[b,c]$, it follows that f is bounded on $[a,b]$ and on $[b,c]$, and hence f is bounded on $[a,c]$. Let $(\{P_n\},\{S_n\})$ be a partition-sample sequence for $[a,c]$. For each n in $\mathbf{Z}^+$ we define a partition P'_n of $[a,b]$ and a partition P''_n of $[b,c]$, and a sample S'_n for P'_n, and a sample S''_n for P''_n as follows:

$$\text{Let } P_n = \{x_0, x_1, \cdots, x_m\}, \;\; S_n = \{s_1, s_2, \cdots, s_m\}.$$

Then there is an index j such that $x_{j-1} \le b \le x_j$.



Let

$$P_n' = \{x_0, \cdots, x_{j-1}, b\}, \quad P_n'' = \{b, x_j, \cdots, x_m\} \tag{8.19}$$

$$S_n' = \{s_1, \cdots, s_{j-1}, b\}, \quad S_n'' = \{b, s_{j+1}, \cdots, s_m\} \tag{8.20}$$

We have

$$\begin{aligned}\sum(f, P_n', S_n') &+ \sum(f, P_n'', S_n'') \\ &= \sum_{i=1}^{j-1} f(s_i)(x_i - x_{i-1}) + f(b)(b - x_{j-1}) + f(b)(x_j - b) \\ &\quad + \sum_{i=j+1}^{m} f(s_i)(x_i - x_{i-1}) \\ &= \sum_{i=1}^{m} f(s_i)(x_i - x_{i-1}) + f(b)(x_j - x_{j-1}) - f(s_j)(x_j - x_{j-1}) \\ &= \sum(f, P_n, S_n) + \Delta_n, \end{aligned} \tag{8.21}$$

where

$$\Delta_n = \Big(f(b) - f(s_j)\Big)(x_j - x_{j-1}).$$

Let M be a bound for f on $[a, c]$. Then

$$|f(b) - f(s_j)| \le |f(b)| + |f(s_j)| \le M + M = 2M.$$

Also,

$$(x_j - x_{j-1}) \le \mu(P_n).$$

Now

$$0 \le |\Delta_n| = |f(b) - f(s_j)| \cdot |x_j - x_{j-1}| \le 2M\mu(P_n).$$

Since

$$\lim\{2M\mu(P_n)\} = 0,$$

it follows from the squeezing rule that $\{|\Delta_n|\} \to 0$ and hence $\{\Delta_n\} \to 0$.

From equation (8.21) we have

$$\sum(f, P_n, S_n) = \sum(f, P_n', S_n') + \sum(f, P_n'', S_n'') - \Delta_n. \tag{8.22}$$

Since $\mu(P_n') \le \mu(P_n)$ and $\mu(P_n'') \le \mu(P_n)$, we see that $(\{P_n'\}, \{S_n'\})$ is a partition-sample sequence on $[a, b]$, and $(\{P_n''\}, \{S_n''\})$ is a partition-sample sequence on $[b, c]$. Since f was given to be integrable on $[a, b]$ and on $[b, c]$, we know that

$$\{\sum(f, P_n', S_n')\} \to \int_a^b f$$

and

$$\{\sum(f, P_n'', S_n'')\} \to \int_b^c f.$$

Hence it follows from (8.22) that

$$\{\sum(f, P_n, S_n)\} \to \int_a^b f + \int_b^c f$$

i.e., f is integrable on $[a, c]$ and

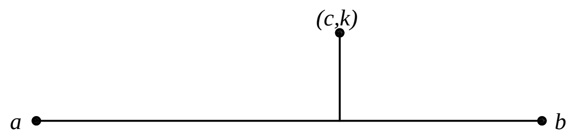
$$\int_a^c f = \int_a^b f + \int_b^c f. \parallel$$

8.23 Corollary. *Let $a_1, a_2, \cdots, a_n$ be real numbers with $a_1 \leq a_2 \cdots \leq a_n$, and let f be a bounded function on $[a_1, a_n]$. If the restriction of f to each of the intervals $[a_1, a_2], [a_2, a_3], \cdots, [a_{n-1}, a_n]$ is integrable, then f is integrable on $[a_1, a_n]$ and*

$$\int_{a_1}^{a_n} f = \int_{a_1}^{a_2} f + \int_{a_2}^{a_3} f + \cdots + \int_{a_{n-1}}^{a_n} f.$$

8.24 Definition (Spike function.) Let $[a, b]$ be an interval. A function $f : [a, b] \to \mathbf{R}$ is called a *spike function*, if there exist numbers c and k, with $c \in [a, b]$ such that

$$f(x) = \begin{cases} 0 & \text{if } x \in [a, b] \setminus \{c\} \\ k & \text{if } x = c. \end{cases}$$



area under graph of spike function

8.25 Theorem (Spike functions are integrable.) *Let a, b, c, k be real numbers with $a < c$ and $a \le b \le c$. Let*

$$f(x) = \begin{cases} 0 & \text{if } x \in [a, c] \setminus \{b\} \\ k & \text{if } x = b. \end{cases}$$

Then f is integrable on $[a, c]$ and $\int_a^c f = 0$.

Proof: Case 1: Suppose $k \ge 0$. Observe that f is increasing on the interval $[a, b]$ and decreasing on the interval $[b, c]$, so f is integrable on each of these intervals. The set of points under the graph of f is the union of a horizontal segment and a vertical segment, and thus is a zero-area set. Hence

$$\int_a^b f = A_a^b f = 0 \qquad \int_b^c f = A_b^c f = 0.$$

By the previous theorem, f is integrable on $[a, c]$, and

$$\int_a^c f = \int_a^b f + \int_b^c f = 0 + 0 = 0$$

.

Case 2: Suppose $k < 0$. Then by case 1 we see that $-f$ is integrable with integral equal to zero, so by the sum theorem for integrals $\int f = 0$ too. |||

8.26 Corollary. *Let a, b, c, k be real numbers with $a < c$ and $a \le b \le c$. Let $f\colon [a, c] \to \mathbf{R}$ be an integrable function and let $g\colon [a, c] \to \mathbf{R}$ be defined by*

$$g(x) = \begin{cases} f(x) & \text{if } x \in [a, c] \setminus \{b\} \\ k & \text{if } x = b. \end{cases}$$

Then g is integrable on $[a, c]$ and $\int_a^c g = \int_a^c f$.

8.27 Corollary. *Let f be an integrable function from an interval $[a, b]$ to $\mathbf{R}$. Let $a_1 \cdots a_n$ be a finite set of distinct points in $\mathbf{R}$, and let $k_1 \cdots k_n$ be a finite set of numbers. Let*

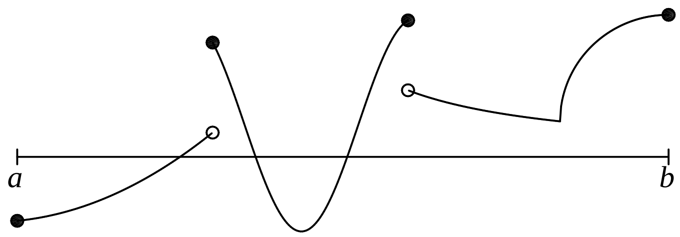
$$g(x) = \begin{cases} f(x) & \text{if } x \in [a, b] \setminus \{a_1, \cdots, a_n\} \\ k_j & \text{if } x = a_j \text{ for some } j \text{ with } 1 \le j \le n. \end{cases}$$

Then g is integrable on $[a, b]$ and $\int_a^b f = \int_a^b g$. Thus we can alter an integrable function on any finite set of points without changing its integrability or its integral.

8.28 Exercise. Prove corollary 8.26, i.e., explain why it follows from theorem 8.25.

8.29 Definition (Piecewise monotonic function.) A function f from an interval $[a, b]$ to $\mathbf{R}$ is *piecewise monotonic* if there are points $a_1, a_2, \cdots, a_n$ in $[a, b]$ with $a < a_1 < a_2 \cdots < a_n < b$ such that f is monotonic on each of the intervals $[a, a_1], [a_1, a_2], \cdots, [a_{n-1}, a_n], [a_n, b]$.

8.30 Example. The function whose graph is sketched below is piecewise monotonic.



piecewise monotonic function

8.31 Theorem. *Every piecewise monotonic function is integrable.*

Proof: This follows from corollary 8.23. |||

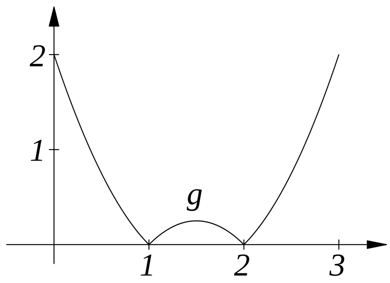
8.32 Exercise. Let

$$f(x) = \begin{cases} x & \text{if } 0 \leq x < 1 \\ x - 1 & \text{if } 1 \leq x \leq 2. \end{cases}$$

Sketch the graph of f. Carefully explain why f is integrable, and find $\int_0^2 f$.

8.33 Example. Let $g(x) = |(x-1)(x-2)|$. Then

$$g(x) = \begin{cases} x^2 - 3x + 2 & \text{for } x \in [0, 1] \\ -x^2 + 3x - 2 & \text{for } x \in [1, 2] \\ x^2 - 3x + 2 & \text{for } x \in [2, 3]. \end{cases}$$



Hence g is integrable on $[0,3]$, and

$$\begin{aligned}\int_0^3 g &= \int_0^1 (x^2-3x+2)dx - \int_1^2 (x^2-3x+2)dx + \int_2^3 (x^2-3x+2)dx \\ &= \left(\frac{1}{3} - 3\cdot\frac{1}{2} + 2\right) - \left(\frac{2^3-1^3}{3} - 3\cdot\frac{2^2-1^2}{2} + 2\right) \\ &\qquad + \left(\frac{3^3-2^3}{3} - 3\cdot\frac{3^2-2^2}{2} + 2\right) \\ &= \left(\frac{1}{3} - \frac{3}{2} + 2\right) - \left(\frac{7}{3} - \frac{9}{2} + 2\right) + \left(\frac{19}{3} - \frac{15}{2} + 2\right) \\ &= \frac{13}{3} + \frac{-9}{2} + 2 = \frac{11}{6}\end{aligned}$$

8.34 Exercise. Calculate the following integrals. Simplify your answers if you can.

a) $\displaystyle\int_0^2 |x^3-1|dx.$

b) $\displaystyle\int_a^b (x-a)(b-x)dx.$ Here $0 < a < b$.

c) $\displaystyle\int_a^b |(x-a)(b-x)|dx.$ Here $0 < a < b$.

d) $\displaystyle\int_0^1 (t^2-2)^3 dt.$

8.3 A Non-integrable Function

We will now give an example of a function that is not integrable. Let

$$S = \{\frac{m}{n} : m \in \mathbf{Z}, n \in \mathbf{Z}^+, \ \ m \text{ and } n \text{ are both odd}\}$$

$$T = \{\frac{m}{n} : m \in \mathbf{Z}, n \in \mathbf{Z}^+, \ \ m \text{ is even and } n \text{ is odd}\}.$$

Then $S \cap T = \emptyset$, since if $\dfrac{m}{n} = \dfrac{p}{q}$ where m, n, and q are odd and p is even, then $mq = np$ which is impossible since mq is odd and np is even.

8.35 Lemma. *Every interval (c, d) in $\mathbf{R}$ with $d - c > 0$ contains a point in S and a point in T.*

Proof: Since $d - c > 0$ we can choose an odd integer n such that $n > \dfrac{3}{d-c}$, i.e., $nd - nc > 3$. Since the interval (nc, nd) has length > 3, it contains at least two integers p, q, say $nc < p < q < nd$. If p and q are both odd, then there is an even integer between them, and if p and q are both even, there is an odd integer between them, so in all cases we can find a set of integers $\{r, s\}$ one of which is even and the other is odd such that $nc < r < s < nd$, i.e., $c < \dfrac{r}{n} < \dfrac{s}{n} < d$. Then $\dfrac{r}{n}$ and $\dfrac{s}{n}$ are two elements of (c, d) one of which is in S, and the other of which is in T. |||

8.36 Example (A non-integrable function.) Let $D: [0, 1] \to \mathbf{R}_{\geq 0}$ be defined by

$$D(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S. \end{cases} \tag{8.37}$$

I will find two partition-sample sequences $(\{P_n\}, \{S_n\})$ and $(\{P_n\}, \{T_n\})$ such that

$$\{\sum(D, P_n, T_n)\} \to 0$$

and

$$\{\sum(D, P_n, S_n)\} \to 1.$$

It then follows that D is not integrable. Let P_n be the regular partition of $[0, 1]$ into n equal subintervals.

$$P_n = \Big\{0, \frac{1}{n}, \frac{2}{n}, \cdots, 1\Big\}.$$

Let S_n be a sample for P_n such that each point in S_n is in S and let T_n be a sample for P_n such that each point in T_n is in T. (We can find such samples by lemma 8.35.) Then for all $n \in \mathbf{Z}^+$

$$\sum(D, P_n, S_n) = \sum_{i=1}^{n} D(s_n)(x_i - x_{i-1}) = 1$$

and

$$\sum(D, P_n, T_n) = \sum_{i=1}^{n} D(t_n)(x_i - x_{i-1}) = 0.$$

So $\lim\{\sum(D, P_n, S_n)\} = 1$ and $\lim\{\sum(D, P_n, T_n)\} = 0$. Both $(\{P_n\}, \{S_n\})$ and $(\{P_n\}, \{T_n\})$ are partition-sample sequences for $[0, 1]$, so it follows that D is not integrable.

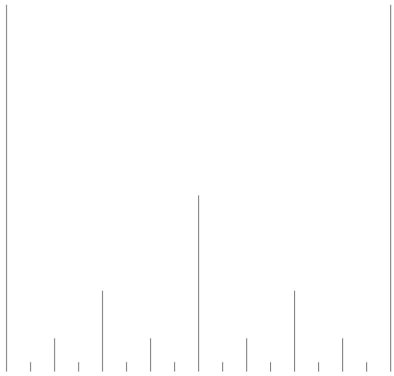
Our example of a non-integrable function is a slightly modified version of an example given by P. G. Lejeune Dirichlet (1805-1859) in 1837. Dirichlet's example was not presented as an example of a non-integrable function (since the definition of integrability in our sense had not yet been given), but rather as an example of how badly behaved a function can be. Before Dirichlet, functions that were this pathological had not been thought of as being functions. It was examples like this that motivated Riemann to define precisely what class of functions are well enough behaved so that we can prove things about them.

8.4 *The Ruler Function

8.38 Example (Ruler function.) We now present an example of an integrable function that is not monotonic on any interval of positive length. Define $R : [0, 1] \to \mathbf{R}$ by

$$R(x) = \begin{cases} 1 & \text{if } x = 0 \text{ or } x = 1 \\ \frac{1}{2^n} & \text{if } x = \frac{q}{2^n} \text{ where } q, n \in \mathbf{Z}^+ \text{ and } q \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

This formula defines $R(x)$ uniquely: If $\dfrac{q}{2^n} = \dfrac{p}{2^m}$ where p and q are odd, then $m = n$. (If $m > n$, we get $2^{m-n}q = p$, which says that an even number is odd.) The set $S_0^1 R$ under the graph of R is shown in the figure.



This set resembles the markings giving fractions of an inch on a ruler, which motivates the name *ruler function* for R. It is easy to see that R is not monotonic on any interval of length > 0. For each $p \in \mathbf{R}$ let $\delta_p \colon \mathbf{R} \to \mathbf{R}$ be defined by

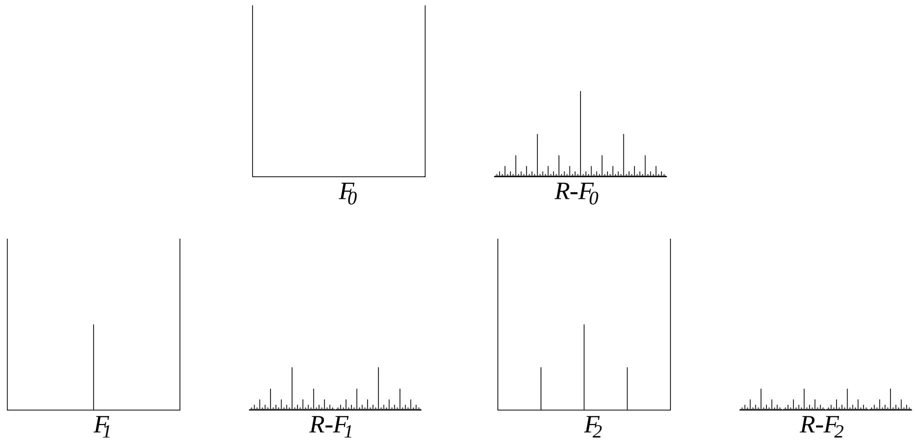
$$\delta_p(t) = \begin{cases} 1 & \text{if } p = t \\ 0 & \text{otherwise.} \end{cases}$$

We have seen that δ_p is integrable on any interval $[a, b]$ and $\int_a^b \delta_p = 0$. Now define a sequence of functions F_j by

$$\begin{aligned} F_0 &= \delta_0 + \delta_1 \\ F_1 &= F_0 + \tfrac{1}{2}\delta_{\frac{1}{2}} \\ F_2 &= F_1 + \tfrac{1}{4}\delta_{\frac{1}{4}} + \tfrac{1}{4}\delta_{\frac{3}{4}} \\ &\vdots \\ F_n &= F_{n-1} + \tfrac{1}{2^n}\sum_{j=1}^{2^{n-1}} \delta_{\frac{2j-1}{2^n}}. \end{aligned}$$

Each function F_j is integrable with integral 0 and

$$|R(x) - F_j(x)| \le \frac{1}{2^{j+1}} \text{ for } 0 \le x \le 1.$$



I will now show that R is integrable.

Let $(\{P_n\}, \{S_n\})$ be a partition-sample sequence for $[0,1]$. I'll show that $\{\sum(R, P_n, S_n)\} \to 0$.

Let ϵ be a generic element in $\mathbf{R}^+$. Observe that if $M \in \mathbf{Z}^+$ then

$$\left(\frac{1}{2^M} < \epsilon\right) \iff \left(M \ln(\frac{1}{2}) < \ln(\epsilon)\right) \iff \left(M > \frac{\ln(\epsilon)}{\ln(\frac{1}{2})}\right).$$

Hence by the Archimedean property, we can choose $M \in \mathbf{Z}^+$ so that $\frac{1}{2^M} < \epsilon$. Then

$$\sum(R, P_n, S_n) = \sum(R - F_M + F_M, P_n, S_n) \tag{8.39}$$

$$= \sum(R - F_M, P_n, S_n) + \sum(F_M, P_n, S_n). \tag{8.40}$$

Now since $0 \leq R(x) - F_M(x) \leq \frac{1}{2^{M+1}} < \frac{1}{2}\epsilon$ for all $x \in [0,1]$, we have

$$\sum(R - F_M, P_n, S_n) \leq \frac{1}{2^{M+1}} < \frac{1}{2}\epsilon \text{ for all } n \in \mathbf{Z}^+.$$

Since F_M is integrable and $\int F_M = 0$, we have $\{\sum(F_M, P_n, S_n)\} \to 0$ so there is an $N \in \mathbf{Z}^+$ such that $|\sum(F_M, P_n, S_n)| < \frac{\epsilon}{2}$ for all $n \in \mathbf{Z}_{\geq N}$. By equation (8.40) we have

$$\begin{aligned} 0 &\leq \sum(R, P_n, S_n) = \sum(R - F_M, P_n, S_n) + \sum(F_M, P_n, S_n) \\ &< \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon \text{ for all } n \in \mathbf{Z}_{\geq N}. \end{aligned}$$

Hence $\{\sum(R, P_n, S_n)\} \to 0$, and hence R is integrable and $\int_0^1 R = 0$.

8.41 Exercise. Let R be the ruler function. We just gave a complicated proof that R is integrable and $\int_0^1 R = 0$. Explain why if you *assume* R is integrable, then it is easy to show that $\int_0^1 R = 0$.

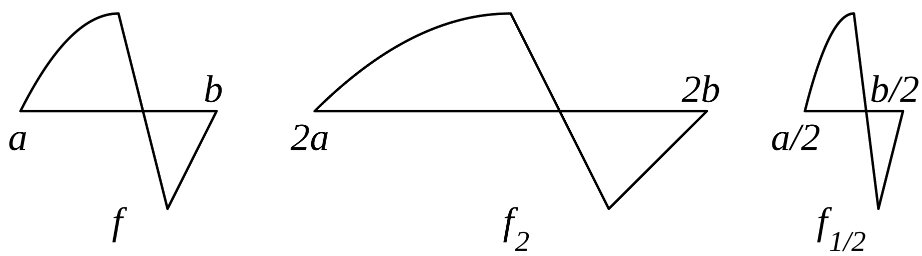
Also show that if you assume that the non-integrable function D in equation (8.37) is integrable then it is easy to show that $\int_0^1 D = 0$.

8.5 Change of Scale

8.42 Definition (Stretch of a function.) Let $[a, b]$ be an interval in $\mathbf{R}$, let $r \in \mathbf{R}^+$, and let $f\colon [a, b] \to \mathbf{R}$. We define a new function $f_r\colon [ra, rb] \to \mathbf{R}$ by

$$f_r(t) = f(\frac{t}{r}) \text{ for all } t \in [ra, rb].$$

If $t \in [ra, rb]$, then $\frac{t}{r} \in [a, b]$, so $f(\frac{t}{r})$ is defined.



The graph of f_r is obtained by stretching the graph of f by a factor of r in the horizontal direction, and leaving it unstretched in the vertical direction. (If $r < 1$ the stretch is actually a shrink.) I will call f_r the *stretch of f by r*.

8.43 Theorem (Change of scale for integrals.) *Let $[a,b]$ be an interval in $\mathbf{R}$ and let $r \in \mathbf{R}^+$. Let $f\colon [a,b] \to \mathbf{R}$ and let f_r be the stretch of f by r. If f is integrable on $[a,b]$ then f_r is integrable on $[ra,rb]$ and $\int_{ra}^{rb} f_r = r\int_a^b f$, i.e.,*

$$\int_{ra}^{rb} f(\frac{x}{r})dx = r\int_a^b f(x)dx. \tag{8.44}$$

Proof: Suppose f is integrable on $[a,b]$. Let $(\{P_n\},\{S_n\})$ be an arbitrary partition-sample sequence for $[ra,rb]$. If

$$P_n = \{x_0,\cdots,x_m\} \text{ and } S_n = \{s_1,\cdots,s_m\},$$

let

$$\frac{1}{r}P_n = \Big\{\frac{x_0}{r},\cdots,\frac{x_m}{r}\Big\} \text{ and } \frac{1}{r}S_n = \Big\{\frac{s_1}{r},\cdots,\frac{s_m}{r}\Big\}.$$

Then $\left(\left\{\frac{1}{r}P_n\right\},\left\{\frac{1}{r}S_n\right\}\right)$ is a partition-sample sequence for $[a,b]$, so $\{\sum(f,\frac{1}{r}P_n,\frac{1}{r}S_n)\} \to \int_a^b f$. Now

$$\begin{aligned}\sum(f_r,P_n,S_n) &= \sum_{i=1}^m f_r(s_i)(x_i - x_{i-1})\\ &= r\sum_{i=1}^m f(\frac{s_i}{r})\Big(\frac{x_i}{r} - \frac{x_{i-1}}{r}\Big) = r\sum\Big(f,\frac{1}{r}P_n,\frac{1}{r}S_n\Big)\end{aligned}$$

so

$$\lim\{\sum(f_r,P_n,S_n)\} = \lim\Big\{r\sum(f,\frac{1}{r}P_n,\frac{1}{r}S_n)\Big\} = r\int_a^b f.$$

This shows that f_r is integrable on $[ra,rb]$, and $\int_{ra}^{rb} f_r = r\int_a^b f$. |||

Remark: The notation f_r is not a standard notation for the stretch of a function, and I will not use this notation in the future. I will usually use the change of scale theorem in the form of equation (8.44), or in the equivalent form

$$\int_A^B g(rx)dx = \frac{1}{r}\int_{rA}^{rB} g(x)dx. \tag{8.45}$$

8.46 Exercise. Explain why formula (8.45) is equivalent to formula (8.44).

8.47 Example. We define π to be the area of the unit circle. Since the unit circle is carried to itself by reflections about the horizontal and vertical axes, we have

$$\pi = 4\ (\text{area (part of unit circle in the first quadrant)}).$$

Since points in the unit circle satisfy $x^2 + y^2 = 1$ or $y^2 = 1 - x^2$, we get

$$\pi = 4\int_0^1 \sqrt{1-x^2}\,dx.$$

We will use this result to calculate the area of a circle of radius a. The points on the circle with radius a and center $\mathbf{0}$ satisfy $x^2 + y^2 = a^2$, and by the same symmetry arguments we just gave

$$\begin{aligned}\text{area(circle of radius } a) &= 4\int_0^a \sqrt{a^2-x^2}\,dx = 4\int_0^a a\sqrt{1-\left(\frac{x}{a}\right)^2}\,dx\\ &= 4a\int_{a\cdot 0}^{a\cdot 1}\sqrt{1-\left(\frac{x}{a}\right)^2}\,dx.\end{aligned}$$

By the change of scale theorem

$$\text{area(circle of radius } a) = 4aa\int_0^1 \sqrt{1-x^2}\,dx = a^2\pi.$$

The formulas

$$\int_0^1 \sqrt{1-x^2}\,dx = \frac{\pi}{4} \text{ and } \int_{-1}^1 \sqrt{1-x^2}\,dx = \frac{\pi}{2}$$

or more generally

$$\int_0^a \sqrt{a^2-x^2}\,dx = \frac{\pi a^2}{4} \text{ and } \int_{-a}^a \sqrt{a^2-x^2} = \frac{\pi a^2}{2},$$

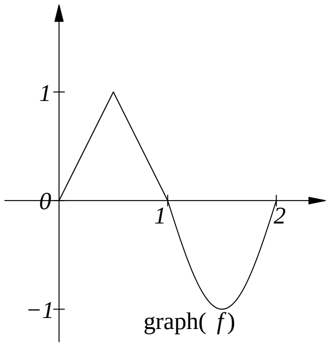
are worth remembering. Actually, these are cases of a formula you already know, since they say that the area of a circle of radius a is πa^2.

8.48 Exercise. Let a, b be positive numbers and let E_{ab} be the set of points inside the ellipse whose equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Calculate the area of E_{ab}.

8.49 Exercise. The figure shows the graph of a function f.



Let functions g, h, k, l, and m be defined by

a) $g(x) = f(\frac{x}{3})$.

b) $h(x) = f(3x)$.

c) $k(x) = f(\frac{x+3}{3})$.

d) $l(x) = f(3x + 3)$.

e) $m(x) = 3f(\frac{x}{3})$.

Sketch the graphs of g,h,k, l, and m on different axes. Use the same scale for all of the graphs, and use the same scale on the x-axis and the y-axis,

8.50 Exercise. The value of $\int_0^1 \frac{1}{1+x^2}\,dx$ is .7854(approximately). Use this fact to calculate approximate values for

$$\int_0^a \frac{1}{a^2+x^2}dx \text{ and } \int_0^{\frac{1}{a}} \frac{1}{1+a^2x^2}dx$$

where $a \in \mathbf{R}^+$. Find numerical values for both of these integrals when $a = \frac{1}{4}$.

8.6 Integrals and Area

8.51 Theorem. *Let f be a piecewise monotonic function from an interval $[a,b]$ to $\mathbf{R}_{\geq 0}$. Then*

$$\int_a^b f = A_a^b f = \alpha(S_a^b f).$$

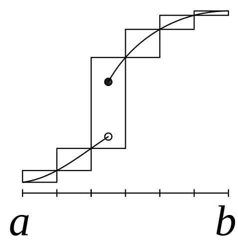
Proof: We already know this result for monotonic functions, and from this the result follows easily for piecewise monotonic functions. |||

Remark Theorem 8.51 is in fact true for all integrable functions from $[a,b]$ to $\mathbf{R}_{\geq 0}$, but the proof is rather technical. Since we will never need the result for functions that are not piecewise monotonic, I will not bother to make an assumption out of it.

8.52 Theorem. *Let $a, b \in \mathbf{R}$ and let $f\colon [a,b] \to \mathbf{R}$ be a piecewise monotonic function. Then the graph of f is a zero-area set.*

Proof: We will show that the theorem holds when f is monotonic on $[a,b]$. It then follows easily that the theorem holds when f is piecewise monotonic on $[a,b]$.

Suppose f is increasing on $[a,b]$. Let $n \in \mathbf{Z}^+$ and let $P = \{x_0, x_1, \cdots, x_n\}$ be the regular partition of $[a,b]$ into n equal subintervals.



Then

$$x_i - x_{i-1} = \frac{b-a}{n} \text{ for } 1 \leq i \leq n$$

and

$$\text{graph}(f) \subset \bigcup_{i=1}^{n} B\Big(x_{i-1}, x_i \colon f(x_{i-1}), f(x_i)\Big).$$

Hence

$$\begin{aligned} 0 \le \alpha\Big(\text{graph}(f)\Big) &\le \alpha\Big(\bigcup_{i=1}^{n} B(x_{i-1}, x_i \colon f(x_{i-1}), f(x_i))\Big) \\ &\le \sum_{i=1}^{n} \alpha\Big(B(x_{i-1}, x_i \colon f(x_{i-1}), f(x_i))\Big) \\ &= \sum_{i=1}^{n} (x_i - x_{i-1})\Big(f(x_i) - f(x_{i-1})\Big) \\ &= \sum_{i=1}^{n} \frac{b-a}{n}\Big(f(x_i) - f(x_{i-1})\Big) \\ &= \frac{b-a}{n} \sum_{i=1}^{n} \Big(f(x_i) - f(x_{i-1})\Big) \\ &= \frac{b-a}{n}\Big(f(b) - f(a)\Big). \end{aligned}$$

Now $\Big\{\dfrac{b-a}{n}\Big(f(b) - f(a)\Big)\Big\} \to 0$, so it follows from the squeezing rule that the constant sequence $\Big\{\alpha\Big(\text{graph}(f)\Big)\Big\}$ converges to 0, and hence

$$\alpha\Big(\text{graph}(f)\Big) = 0. \quad \|$$

Remark: Theorem 8.52 is actually valid for all integrable functions on $[a, b]$.

8.53 Theorem (Area between graphs.) *Let f, g be piecewise monotonic functions on an interval $[a, b]$ such that $g(x) \le f(x)$ for all $x \in [a, b]$. Let*

$$S = \{(x, y) \colon a \le x \le b \text{ and } g(x) \le y \le f(x)\}.$$

Then

$$\text{area}(S) = \int_a^b f(x) - g(x)\, dx.$$

Proof: Let M be a lower bound for g, so that

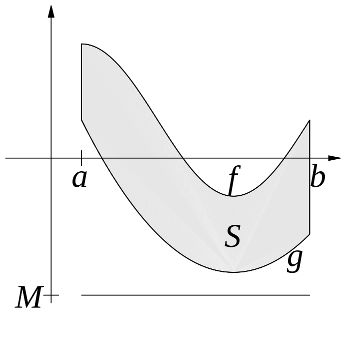
$$0 \le g(x) - M \le f(x) - M \text{ for all } x \in [a, b].$$

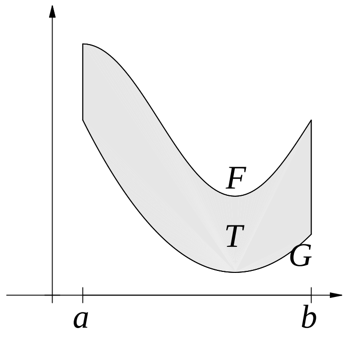
Let

$$F(x) = f(x) - M, \qquad G(x) = g(x) - M$$

for all $x \in [a, b]$, and let

$$T = \{(x, y) \colon a \le x \le b \text{ and } G(x) \le y \le F(x)\}.$$





Then

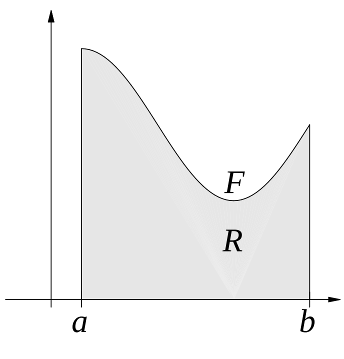
$$
\begin{aligned}
(x,y)\in T &\iff a\le x\le b \text{ and } G(x)\le y\le F(x)\\
&\iff a\le x\le b \text{ and } g(x)-M\le y\le f(x)-M\\
&\iff a\le x\le b \text{ and } g(x)\le y+M\le f(x)\\
&\iff (x,y+M)\in S\\
&\iff (x,y)+(0,M)\in S.
\end{aligned}
$$

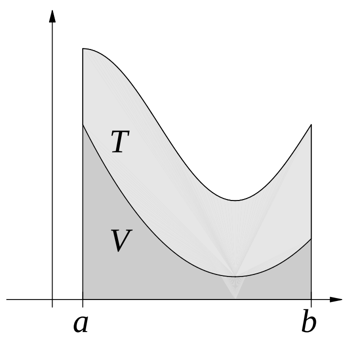
It follows from translation invariance of area that

$$\text{area}(S)=\text{area}(T).$$

Let

$$
\begin{aligned}
R &= \{(x,y)\colon a\le x\le b \text{ and } 0\le y\le F(x)\} = S_a^b F,\\
V &= \{(x,y)\colon a\le x\le b \text{ and } 0\le y\le G(x)\} = S_a^b G.
\end{aligned}
$$





Then $V\cup T=R$, and

$$V\cap T=\{(x,y)\colon a\le x\le b \text{ and } y=G(x)\}=\text{graph}(G).$$

It follows from theorem 8.52 that V and T are almost disjoint, so

$$\text{area}(R) = \text{area}(V \cup T) = \text{area}(V) + \text{area}(T),$$

and thus

$$\text{area}(T) = \text{area}(R) - \text{area}(V).$$

By theorem 8.51 we have

$$\text{area}(R) = \text{area}(S_a^b F) = \int_a^b F(x)\, dx$$

and

$$\text{area}(V) = \text{area}(S_a^b G) = \int_a^b G(x)\, dx.$$

Thus

$$\begin{aligned}
\text{area}(S) &= \text{area}(T) = \text{area}(R) - \text{area}(V) \\
&= \int_a^b F(x)\, dx - \int_a^b G(x)\, dx \\
&= \int_a^b \big(F(x) - G(x)\big)\, dx \\
&= \int_a^b \Big(f(x) - M - \big(g(x) - M\big)\Big)\, dx \\
&= \int_a^b f(x) - g(x)\, dx. \ \|
\end{aligned}$$

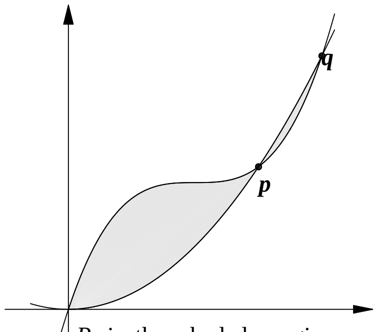
Remark: Theorem 8.53 is valid for all integrable functions f and g. This follows from our proof and the fact that theorems 8.51 and 8.52 are both valid for all integrable functions.

8.54 Example. We will find the area of the set R in the figure, which is bounded by the graphs of f and g where

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

and

$$g(x) = x^3 - 3x^2 + 3x.$$



Now

$$g(x) - f(x) = x^3 - 3x^2 + 3x - \frac{1}{2}x^2 = x^3 - \frac{7}{2}x^2 + 3x$$
$$= x(x^2 - \frac{7}{2}x + 3) = x(x-2)(x - \frac{3}{2}).$$

Hence

$$(g(x) = f(x)) \iff (g(x) - f(x) = 0) \iff \left(x \in \left\{0, \frac{3}{2}, 2\right\}\right).$$

It follows that the points $\mathbf{p}$ and $\mathbf{q}$ in the figure are

$$\mathbf{p} = (\frac{3}{2}, f(\frac{3}{2})) = (\frac{3}{2}, \frac{9}{8}) \text{ and } \mathbf{q} = (2, f(2)) = (2, 2).$$

Also, since $x(x-2) \leq 0$ for all $x \in [0, 2]$,

$$g(x) - f(x) \geq 0 \iff x - \frac{3}{2} \leq 0 \iff x \leq \frac{3}{2}.$$

(This is clear from the picture, assuming that the picture is accurate.) Thus

$$\text{area}(R) = \int_0^{\frac{3}{2}} (g - f) + \int_{\frac{3}{2}}^2 (f - g)$$
$$= \int_0^{\frac{3}{2}} (x^3 - \frac{7}{2}x^2 + 3x)dx - \int_{\frac{3}{2}}^2 (x^3 - \frac{7}{2}x^2 + 3x)dx$$
$$= \left(\frac{(\frac{3}{2})^4 - 0^4}{4}\right) - \frac{7}{2}\left(\frac{(\frac{3}{2})^3 - 0^3}{3}\right) + 3\left(\frac{(\frac{3}{2})^2 - 0^2}{2}\right)$$
$$- \left(\frac{2^4 - (\frac{3}{2})^4}{4}\right) + \frac{7}{2}\left(\frac{2^3 - (\frac{3}{2})^3}{3}\right) - 3\left(\frac{2^2 - (\frac{3}{2})^2}{2}\right).$$

We have now found the area, but the answer is not in a very informative form. It is not clear whether the number we have found is positive. It would be reasonable to use a calculator to simplify the result, but my experience with calculators is that I am more likely to make an error entering this into my calculator than I am to make an error by doing the calculation myself, so I will continue. I notice that three terms in the answer are repeated twice, so I have

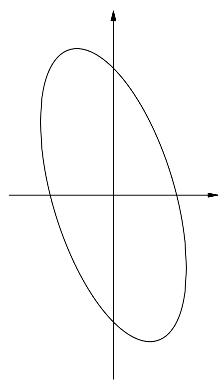
$$\begin{aligned} \text{area}(R) &= 2\left(\frac{81}{64} - \frac{63}{16} + \frac{27}{8}\right) - 4 + \frac{28}{3} - 6 \\ &= \frac{81}{32} - \frac{63}{8} + \frac{27}{4} - \frac{2}{3} \\ &= (2 + \frac{17}{32}) - (8 - \frac{1}{8}) + (6 + \frac{3}{4}) - \frac{2}{3} \\ &= \frac{17}{32} + \frac{1}{8} + \frac{3}{4} - \frac{2}{3} = \frac{21}{32} + \frac{1}{12} = \frac{63+8}{96} = \frac{71}{96}. \end{aligned}$$

Thus the area is about 0.7 From the sketch I expect the area to be a little bit smaller than 1, so the answer is plausible.

8.55 Exercise. The curve whose equation is

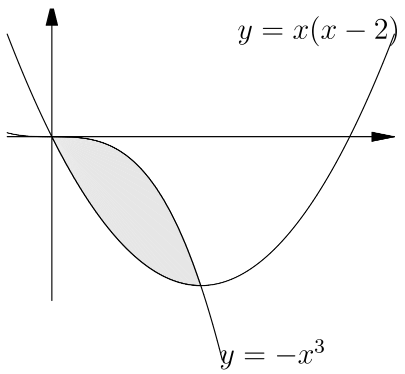
$$y^2 + 2xy + 2x^2 = 4 \tag{8.56}$$

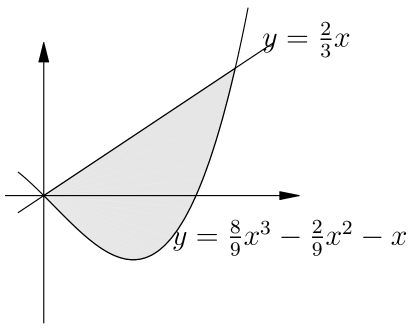
is shown in the figure. Find the area enclosed by the curve.



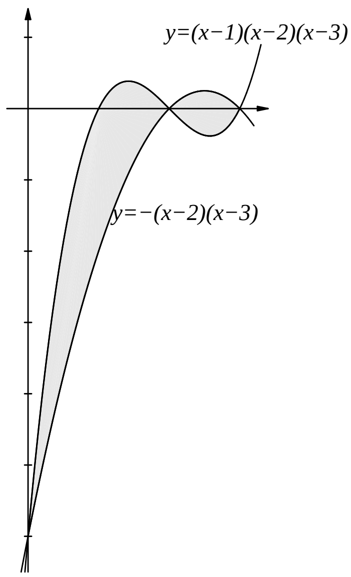
(The set whose area we want to find is bounded by the graphs of the two functions. You can find the functions by considering equation (8.56) as a quadratic equation in y and solving for y as a function of x.)

8.57 Exercise. Find the areas of the two sets shaded in the figures below:





8.58 Exercise. Find the area of the shaded region.



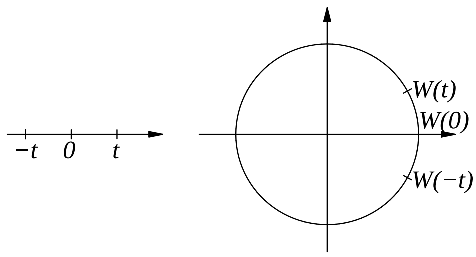
Chapter 9

Trigonometric Functions

9.1 Properties of Sine and Cosine

9.1 Definition ($W(t)$.) We define a function $W: \mathbf{R} \to \mathbf{R}^2$ as follows.

If $t \geq 0$, then $W(t)$ is the point on the unit circle such that the length of the arc joining $(1,0)$ to $W(t)$ (measured in the counterclockwise direction) is equal to t. (There is an optical illusion in the figure. The length of segment $[0,t]$ is equal to the length of arc $W(0)W(t)$.)



Thus to find $W(t)$, you should start at $(1,0)$ and move along the circle in a counterclockwise direction until you've traveled a distance t. Since the circumference of the circle is 2π, we see that $W(2\pi) = W(4\pi) = W(0) = (1,0)$. (Here we assume Archimedes' result that the area of a circle is half the circumference times the radius.) If $t < 0$, we define

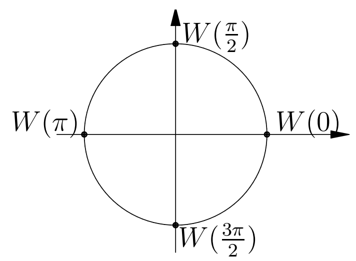
$$W(t) = H(W(-t)) \text{ for } t < 0 \tag{9.2}$$

where H is the reflection about the horizontal axis. Thus if $t < 0$, then $W(t)$ is the point obtained by starting at $(1,0)$ and moving $|t|$ along the unit circle in the clockwise direction.

Remark: The definition of W depends on several ideas that we have not defined or stated assumptions about, e.g., *length of an arc* and *counterclockwise direction*. I believe that the amount of work it takes to formalize these ideas at this point is not worth the effort, so I hope your geometrical intuition will carry you through this chapter. (In this chapter we will assume quite a bit of Euclidean geometry, and a few properties of area that do not follow from our assumptions stated in chapter 5.)

A more self contained treatment of the trigonometric functions can be found in [44, chapter 15], but the treatment given there uses ideas that we will consider later, (e.g. derivatives, inverse functions, the intermediate value theorem, and the fundamental theorem of calculus) in order to *define* the trigonometric functions.

We have the following values for W:



$$W(0) = (1,0) \tag{9.3}$$
$$W\Big(\frac{\pi}{2}\Big) = (0,1) \tag{9.4}$$
$$W(\pi) = (-1,0) \tag{9.5}$$
$$W\Big(\frac{3\pi}{2}\Big) = (0,-1) \tag{9.6}$$
$$W(2\pi) = (1,0) = W(0). \tag{9.7}$$

In general

$$W(t+2\pi k) = W(t) \text{ for all } t \in \mathbf{R} \text{ and all } k \in \mathbf{Z}. \tag{9.8}$$

9.9 Definition (Sine and cosine.) In terms of coordinates, we write

$$W(t) = \Big(\cos(t), \sin(t)\Big).$$

(We read "$\cos(t)$" as "cosine of t", and we read "$\sin(t)$" as "sine of t".)

Since $W(t)$ is on the unit circle, we have

$$\sin^2(t) + \cos^2(t) = 1 \text{ for all } t \in \mathbf{R},$$

and

$$-1 \leq \sin t \leq 1, \quad -1 \leq \cos t \leq 1 \text{ for all } t \in \mathbf{R}.$$

The equations (9.3) - (9.8) show that

$$\begin{array}{ll} \cos(0) = 1, & \sin(0) = 0, \\ \cos\left(\frac{\pi}{2}\right) = 0, & \sin\left(\frac{\pi}{2}\right) = 1, \\ \cos(\pi) = -1, & \sin(\pi) = 0, \\ \cos\left(\frac{3\pi}{2}\right) = 0, & \sin\left(\frac{3\pi}{2}\right) = -1, \end{array}$$

and

$$\begin{aligned} \cos(t + 2\pi k) = \cos t & \quad \text{for all } t \in \mathbf{R} \text{ and all } k \in \mathbf{Z}, \\ \sin(t + 2\pi k) = \sin t & \quad \text{for all } t \in \mathbf{R} \text{ and all } k \in \mathbf{Z}. \end{aligned}$$

In equation (9.2) we defined

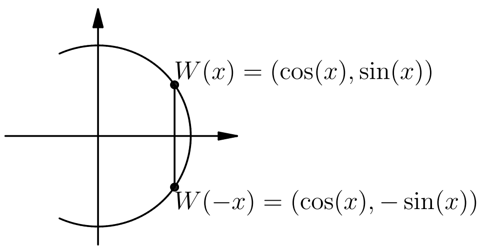
$$W(t) = H(W(-t)) \text{ for } t < 0.$$

Thus for $t < 0$,

$$W(-t) = H(H(W(-t))) = H(W(t)) = H(W(-(-t))),$$

and it follows that

$$W(t) = H(W(-t)) \text{ for all } t \in \mathbf{R}.$$



In terms of components

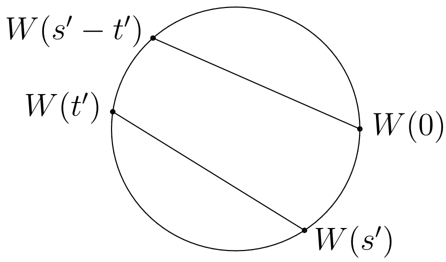
$$\begin{aligned}\Big(\cos(-t), \sin(-t)\Big) &= W(-t) = H(W(t)) = H(\cos(t), \sin(t)) \\ &= \Big(\cos(t), -\sin(t)\Big)\end{aligned}$$

and consequently

$$\cos(-t) = \cos(t) \text{ and } \sin(-t) = -\sin(t) \text{ for all } t \in \mathbf{R}.$$

Let s, t be arbitrary real numbers. Then there exist integers k and l such that $s + 2\pi k \in [0, 2\pi)$ and $t + 2\pi l \in [0, 2\pi)$. Let

$$s' = s + 2\pi k \text{ and } t' = t + 2\pi l.$$



Then $s' - t' = (s - t) + 2\pi(k - l)$, so

$$W(s) = W(s'),\ W(t) = W(t'),\ W(s - t) = W(s' - t').$$

Suppose $t' \le s'$ (see figure). Then the length of the arc joining $W(s')$ to $W(t')$ is $s' - t'$ which is the same as the length of the arc joining $(1, 0)$ to $W(s' - t')$. Since equal arcs in a circle subtend equal chords, we have

$$\text{dist}\Big(W(s'), W(t')\Big) = \text{dist}\Big(W(s' - t'), (1, 0)\Big)$$

and hence

$$\text{dist}\Big(W(s), W(t)\Big) = \text{dist}\Big(W(s - t), (1, 0)\Big). \tag{9.10}$$

You can verify that this same relation holds when $s' < t'$.

9.11 Theorem (Addition laws for sine and cosine.) *For all real numbers s and t,*

$$\cos(s + t) = \cos(s)\cos(t) - \sin(s)\sin(t) \tag{9.12}$$
$$\cos(s - t) = \cos(s)\cos(t) + \sin(s)\sin(t) \tag{9.13}$$
$$\sin(s + t) = \sin(s)\cos(t) + \cos(s)\sin(t) \tag{9.14}$$
$$\sin(s - t) = \sin(s)\cos(t) - \cos(s)\sin(t). \tag{9.15}$$

Proof: From (9.10) we know

$$\operatorname{dist}\Big(W(s), W(t)\Big) = \operatorname{dist}\Big(W(s-t), (1,0)\Big),$$

i.e.,

$$\operatorname{dist}\Big((\cos(s), \sin(s)), (\cos(t), \sin(t))\Big) = \operatorname{dist}\Big((\cos(s-t), \sin(s-t)), (1,0)\Big).$$

Hence

$$\Big(\cos(s) - \cos(t)\Big)^2 + \Big(\sin(s) - \sin(t)\Big)^2 = \Big(\cos(s-t) - 1\Big)^2 + \Big(\sin(s-t)\Big)^2.$$

By expanding the squares and using the fact that $\sin^2(u) + \cos^2(u) = 1$ for all u, we conclude that

$$\cos(s)\cos(t) + \sin(s)\sin(t) = \cos(s-t). \tag{9.16}$$

This is equation (9.13). To get equation (9.12) replace t by $-t$ in (9.16). If we take $s = \frac{\pi}{2}$ in equation (9.16) we get

$$\cos\Big(\frac{\pi}{2}\Big)\cos(t) + \sin\Big(\frac{\pi}{2}\Big)\sin(t) = \cos\Big(\frac{\pi}{2} - t\Big)$$

or

$$\sin(t) = \cos\Big(\frac{\pi}{2} - t\Big) \text{ for all } t \in \mathbf{R}.$$

If we replace t by $\Big(\frac{\pi}{2} - t\Big)$ in this equation we get

$$\sin\Big(\frac{\pi}{2} - t\Big) = \cos\Big(\frac{\pi}{2} - (\frac{\pi}{2} - t)\Big) = \cos t \text{ for all } t \in \mathbf{R}.$$

Now in equation (9.16) replace s by $\frac{\pi}{2} - s$ and get

$$\cos\Big(\frac{\pi}{2} - s\Big)\cos(t) + \sin\Big(\frac{\pi}{2} - s\Big)\sin(t) = \cos\Big(\frac{\pi}{2} - s - t\Big)$$

or

$$\sin s \cos t + \cos s \sin t = \sin(s+t),$$

which is equation (9.14). Finally replace t by $-t$ in this last equation to get (9.15). |||

In the process of proving the last theorem we proved the following:

9.17 Theorem (Reflection law for sin and cos.) *For all $x \in \mathbf{R}$,*

$$\cos(x) = \sin(\frac{\pi}{2} - x) \text{ and } \sin(x) = \cos(\frac{\pi}{2} - x). \tag{9.18}$$

9.19 Theorem (Double angle and half angle formulas.) *For all $t \in \mathbf{R}$ we have*

$$\begin{aligned}
&\sin(2t) = 2\sin t \cos t,\\
&\cos(2t) = \cos^2 t - \sin^2 t = 2\cos^2 t - 1 = 1 - 2\sin^2 t,\\
&\sin^2\Big(\frac{t}{2}\Big) = \frac{1-\cos t}{2},\\
&\cos^2\Big(\frac{t}{2}\Big) = \frac{1+\cos t}{2}.
\end{aligned}$$

9.20 Exercise. Prove the four formulas stated in theorem 9.19.

9.21 Theorem (Products and differences of sin and cos.) *For all s, t in $\mathbf{R}$,*

$$\cos(s)\cos(t) = \frac{1}{2}[\cos(s-t) + \cos(s+t)], \tag{9.22}$$

$$\cos(s)\sin(t) = \frac{1}{2}[\sin(s+t) - \sin(s-t)], \tag{9.23}$$

$$\sin(s)\sin(t) = \frac{1}{2}[\cos(s-t) - \cos(s+t)], \tag{9.24}$$

$$\cos(s) - \cos(t) = -2\sin\Big(\frac{s+t}{2}\Big)\sin\Big(\frac{s-t}{2}\Big), \tag{9.25}$$

$$\sin(s) - \sin(t) = 2\cos\Big(\frac{s+t}{2}\Big)\sin\Big(\frac{s-t}{2}\Big). \tag{9.26}$$

Proof: We have

$$\cos(s+t) = \cos(s)\cos(t) - \sin(s)\sin(t)$$

and

$$\cos(s-t) = \cos(s)\cos(t) + \sin(s)\sin(t).$$

By adding these equations, we get (9.22). By subtracting the first from the second, we get (9.24).

In equation (9.24) replace s by $\dfrac{s+t}{2}$ and replace t by $\dfrac{t-s}{2}$ to get

$$\sin\Big(\frac{s+t}{2}\Big)\sin\Big(\frac{t-s}{2}\Big) = \frac{1}{2}\Big[\cos\Big(\frac{s+t}{2} - \frac{t-s}{2}\Big) - \cos\Big(\frac{s+t}{2} + \frac{t-s}{2}\Big)\Big]$$

or

$$-\sin\Big(\frac{s+t}{2}\Big)\sin\Big(\frac{s-t}{2}\Big) = \frac{1}{2}[\cos(s) - \cos(t)].$$

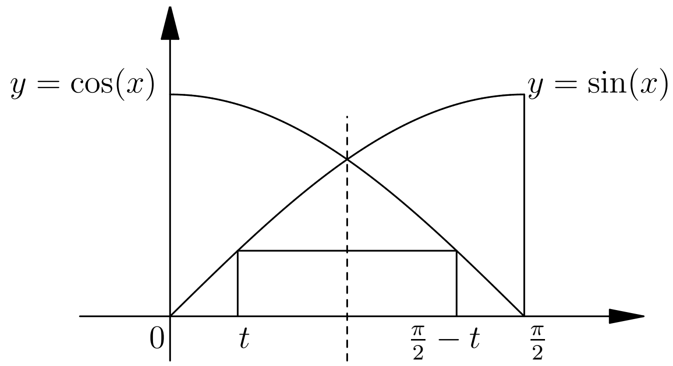
This yields equation (9.25).

9.27 Exercise. Prove equations (9.23) and (9.26).

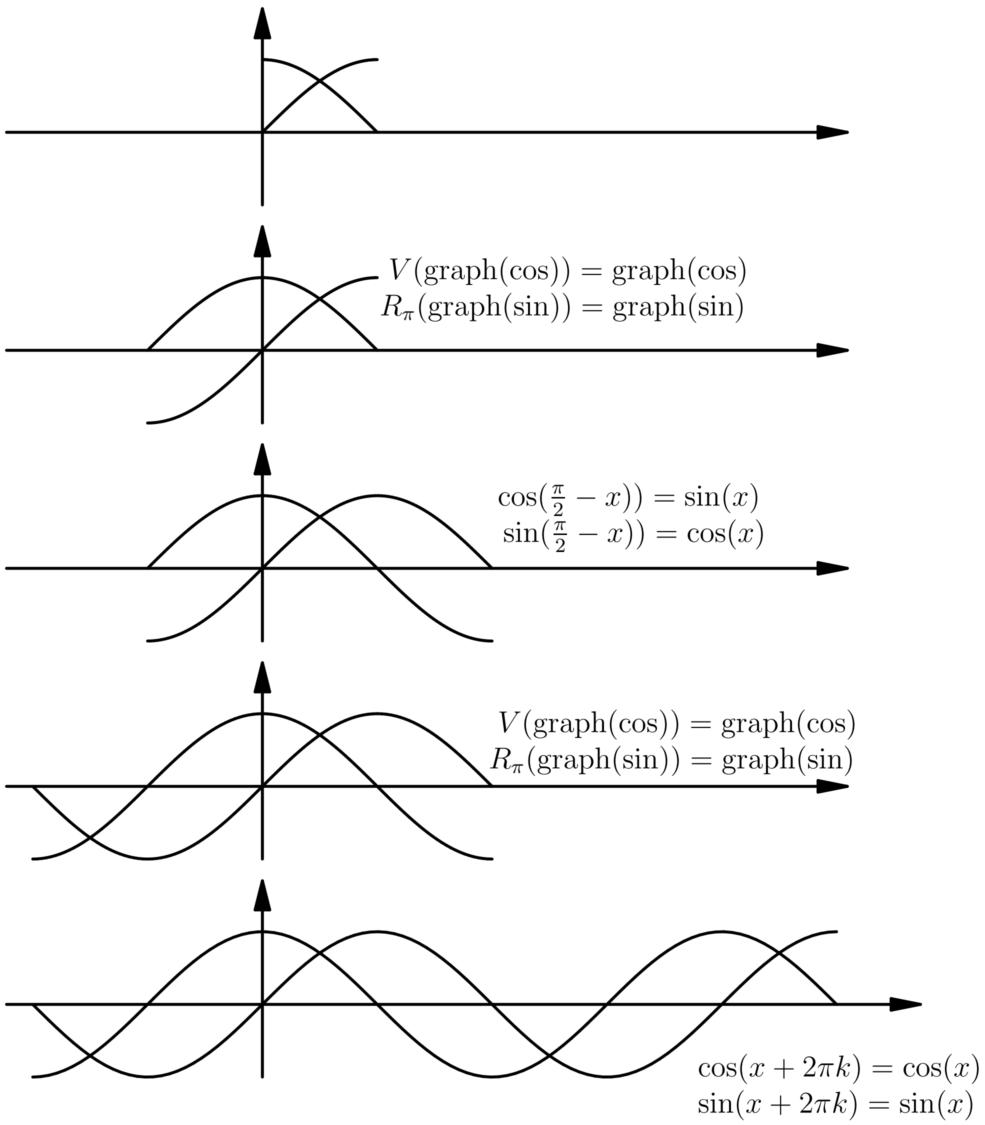
From the geometrical description of sine and cosine, it follows that as t increases for 0 to $\dfrac{\pi}{2}$, $\sin(t)$ increases from 0 to 1 and $\cos(t)$ decreases from 1 to 0. The identities

$$\sin\Big(\frac{\pi}{2} - t\Big) = \cos(t) \text{ and } \cos\Big(\frac{\pi}{2} - t\Big) = \sin(t)$$

indicate that a reflection about the vertical line through $x = \dfrac{\pi}{4}$ carries the graph of sin onto the graph of cos, and vice versa.



$\cos(\frac{\pi}{2} - t) = \sin(t) \qquad \sin(\frac{\pi}{2} - t) = \cos(t)$



The condition $\cos(-x) = \cos x$ indicates that the reflection about the vertical axis carries the graph of cos to itself.

The relation $\sin(-x) = -\sin(x)$ shows that

$$\begin{aligned}(x, y) \in \text{graph}(\sin) &\implies y = \sin(x)\\ &\implies -y = -\sin(x) = \sin(-x)\\ &\implies (-x, -y) = (-x, \sin(-x))\\ &\implies (-x, -y) \in \text{graph}(\sin)\\ &\implies R_\pi(x, y) \in \text{graph}(\sin)\end{aligned}$$

i.e., the graph of sin is carried onto itself by a rotation through π about the origin.
We have

$$\sin\left(\frac{\pi}{4}\right) = \cos\left(\frac{\pi}{2} - \frac{\pi}{4}\right) = \cos\left(\frac{\pi}{4}\right)$$

and $1 = \sin^2\left(\frac{\pi}{4}\right) + \cos^2\left(\frac{\pi}{4}\right) = 2\cos^2\left(\frac{\pi}{4}\right)$, so $\cos^2\left(\frac{\pi}{4}\right) = \frac{1}{2}$ and $\sin\frac{\pi}{4} = \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2} = .707$ (approximately).

With this information we can make a reasonable sketch of the graph of sin and cos (see page 197).

9.28 Exercise. Show that

$$\cos(3x) = 4\cos^3(x) - 3\cos(x) \text{ for all } x \in \mathbf{R}.$$

9.29 Exercise. Complete the following table of sines and cosines:

	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π
sin	0		$\frac{\sqrt{2}}{2}$		1				0
cos	1		$\frac{\sqrt{2}}{2}$		0				-1

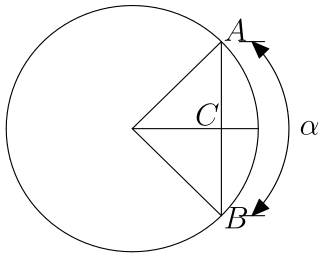
	π	$\frac{7\pi}{6}$	$\frac{5\pi}{4}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{7\pi}{4}$	$\frac{11\pi}{6}$	2π
sin	0				-1				0
cos	-1				0				1

$$\frac{\sqrt{2}}{2} = .707$$

Include an explanation for how you found $\sin\frac{\pi}{6}$ and $\cos\frac{\pi}{6}$ (or $\sin\frac{\pi}{3}$ and $\cos\frac{\pi}{3}$).

For the remaining values you do not need to include an explanation.

Most of the material from this section was discussed by Claudius Ptolemy (fl. 127-151 AD). The functions considered by Ptolemy were not the sine and cosine, but the *chord*, where the chord of an arc α is the length of the segment joining its endpoints.



$AB = \text{chord}(\alpha)$ $\quad AC = \sin(\frac{\alpha}{2})$

$$\text{chord }(\alpha) = 2\sin(\frac{\alpha}{2}). \tag{9.30}$$

Ptolemy's chords are functions of arcs (measured in degrees), not of numbers. Ptolemy's addition law for sin was (roughly)

$$D \cdot \text{ chord}(\beta - \alpha) = \text{chord}(\beta)\text{chord}(180^\circ - \alpha) - \text{chord}(180^\circ - \beta)\text{chord}(\alpha),$$

where D is the diameter of the circle, and $0^\circ < \alpha < \beta < 180^\circ$. Ptolemy produced tables equivalent to tables of $\sin(\alpha)$ for $\left(\frac{1}{4}\right)^\circ \le \alpha \le 90^\circ$ in intervals of $\left(\frac{1}{4}\right)^\circ$. All calculations were made to 3 sexagesimal (base 60) places.

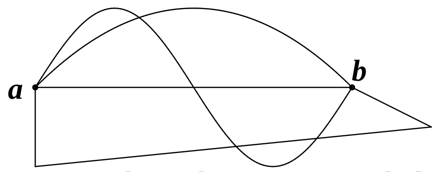
The etymology of the word *sine* is rather curious[42, pp 615-616]. The function we call sine was first given a name by Āryabhaṭa near the start of the sixth century AD. The name meant "half chord" and was later shortened to *jyā* meaning "chord". The Hindu word was translated into Arabic as *jîba*, which was a meaningless word phonetically derived from *jyā*, but (because the vowels in Arabic were not written) was written the same as *jaib*, which means bosom. When the Arabic was translated into Latin it became *sinus*. (*Jaib* means bosom, bay, or breast: *sinus* means bosom, bay or the fold of a toga around the breast.) The English word *sine* is derived from *sinus* phonetically.

9.31 Entertainment (Calculation of sines.) Design a computer program that will take as input a number x between 0 and .5, and will calculate $\sin(\pi x)$. (I choose $\sin(\pi x)$ instead of $\sin(x)$ so that you do *not* need to know the value of π to do this.)

9.2 Calculation of π

The proof of the next lemma depends on the following assumption, which is explicitly stated by Archimedes [2, page 3]. This assumption involves the ideas of *curve with given endpoints* and *length of curve* (which I will leave undiscussed).

9.32 Assumption. Let **a** and **b** be points in $\mathbf{R}^2$. Then of all curves with endpoints **a** and **b**, the segment [**ab**] is the shortest.



curves with endpoints ***a*** and ***b***

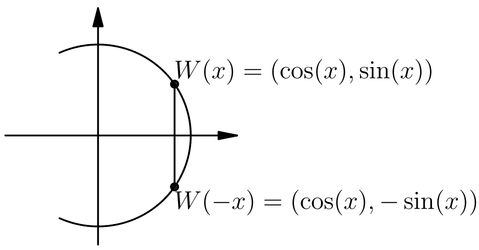
9.33 Lemma.

$$\sin(x) < x \textit{ for all } x \in \mathbf{R}^+,$$

and

$$|\sin(x)| \leq |x| \textit{ for all } x \in \mathbf{R}.$$

Proof:
Case 1: Suppose $0 < x < \dfrac{\pi}{2}$.



Then (see the figure) the length of the arc joining $W(-x)$ to $W(x)$ in the first and fourth quadrants is $x + x = 2x$. (This follows from the definition of W.) The length of the segment $[W(x)W(-x)]$ is $2\sin(x)$. By our assumption, $2\sin(x) \leq 2x$, i.e., $\sin(x) \leq x$. Since both x and $\sin(x)$ are positive when $0 < x < \frac{\pi}{2}$, we also have $|\sin(x)| \leq |x|$.
Case 2: Suppose $x \geq \dfrac{\pi}{2}$. Then

$$\sin(x) \leq |\sin(x)| \leq 1 < \frac{\pi}{2} \leq x = |x|$$

so $\sin(x) \leq x$ and $|\sin(x)| \leq |x|$ in this case also. This proves the first assertion of lemma 9.33. If $x < 0$, then $-x > 0$, so

$$|\sin(x)| = |\sin(-x)| \leq |-x| = |x|.$$

Thus

$$|\sin(x)| \leq |x| \text{ for all } x \in \mathbf{R} \setminus \{0\},$$

and since the relation clearly holds when $x = 0$ the lemma is proved. |||

9.34 Lemma (Limits of sine and cosine.) *Let $a \in \mathbf{R}$. Let $\{a_n\}$ be a sequence in $\mathbf{R}$ such that $\{a_n\} \to a$. Then*

$$\{\cos(a_n)\} \to \cos(a) \text{ and } \{\sin(a_n)\} \to \sin(a).$$

Proof: By (9.25) we have

$$\cos(a_n) - \cos(a) = -2\sin\Big(\frac{a_n + a}{2}\Big)\sin\Big(\frac{a_n - a}{2}\Big),$$

so

$$\begin{aligned} 0 \leq |\cos(a_n) - \cos(a)| &\leq 2\Big|\sin\Big(\frac{a_n + a}{2}\Big)\Big|\Big|\sin\Big(\frac{a_n - a}{2}\Big)\Big| \\ &\leq 2\Big|\sin\Big(\frac{a_n - a}{2}\Big)\Big| \leq 2\Big|\frac{a_n - a}{2}\Big| = |a_n - a|. \end{aligned}$$

If $\{a_n\} \to a$, then $\{|a_n - a|\} \to 0$, so by the squeezing rule,

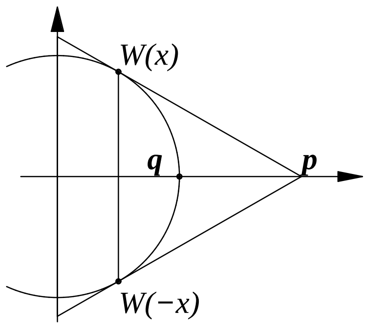
$$\{|\cos(a_n) - \cos(a)|\} \to 0.$$

This means that $\{\cos(a_n)\} \to \cos(a)$.

The proof that $\{\sin(a_n)\} \to \sin(a)$ is similar. |||

The proof of the next lemma involves another new assumption.

9.35 Assumption. Suppose $0 < x < \frac{\pi}{2}$. Let the tangent to the unit circle at $W(x)$ intersect the x axis at $\mathbf{p}$, and let $\mathbf{q} = (1, 0)$.



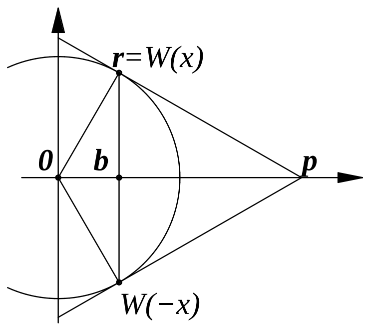
Then the circular arc joining $W(x)$ to $W(-x)$ (and passing through $\mathbf{q}$) is shorter than the curve made of the two segments $[W(x)\mathbf{p}]$ and $[\mathbf{p}W(-x)]$ (see the figure).

Remark: Archimedes makes a general assumption about curves that are *concave in the same direction* [2, pages 2-4] which allows him to prove our assumption.

9.36 Lemma. *If* $0 < x < \frac{\pi}{2}$, *then*

$$x \leq \frac{\sin(x)}{\cos(x)}.$$

Proof: Suppose $0 < x < \frac{\pi}{2}$. Draw the tangents to the unit circle at $W(x)$ and $W(-x)$ and let the point at which they intersect the x-axis be $\mathbf{p}$. (By symmetry both tangents intersect the x-axis at the same point.) Let $\mathbf{b}$ be the point where the segment $[W(x)W(-x)]$ intersects the x-axis, and let $\mathbf{r} = W(x)$. Triangles $\mathbf{0br}$ and $\mathbf{0rp}$ are similar since they are right triangles with a common acute angle.



Hence

$$\frac{\text{distance}(\mathbf{r},\mathbf{b})}{\text{distance}(\mathbf{0},\mathbf{b})} = \frac{\text{distance}(\mathbf{p},\mathbf{r})}{\text{distance}(\mathbf{0},\mathbf{r})}$$

i.e.,

$$\frac{\sin(x)}{\cos(x)} = \frac{\text{distance}(\mathbf{p},\mathbf{r})}{1}.$$

Now the length of the arc joining $W(x)$ to $W(-x)$ is $2x$, and the length of the broken line from $\mathbf{r}$ to $\mathbf{p}$ to $W(-x)$ is $2\Big(\text{distance}(\mathbf{p},\mathbf{r})\Big) = 2\dfrac{\sin(x)}{\cos(x)}$, so by assumption 9.35,

$$2x \leq 2\frac{\sin(x)}{\cos x}$$

i.e.,

$$x \leq \frac{\sin(x)}{\cos(x)}.$$

This proves our lemma. ‖

9.37 Theorem. *Let $\{x_n\}$ be any sequence such that $x_n \neq 0$ for all n, and $\{x_n\} \to 0$. Then*

$$\Big\{\frac{\sin(x_n)}{x_n}\Big\} \to 1. \tag{9.38}$$

Hence if $\sin(x_n) \neq 0$ *for all* $n \in \mathbf{Z}^+$ *we also have*

$$\Big\{\frac{x_n}{\sin(x_n)}\Big\} \to 1.$$

Proof: If $x \in (0, \frac{\pi}{2})$, then it follows from lemma(9.36) that $\cos(x) \leq \dfrac{\sin(x)}{x}$. Since

$$\cos(-x) = \cos(x) \text{ and } \frac{\sin(-x)}{-x} = \frac{\sin(x)}{x},$$

it follows that

$$\cos(x) \leq \frac{\sin(x)}{x} \text{ whenever } 0 < |x| < \frac{\pi}{2}.$$

Hence by lemma 9.33 we have

$$\cos(x) \leq \frac{\sin(x)}{x} \leq 1 \text{ whenever } 0 < |x| < \frac{\pi}{2}. \tag{9.39}$$

Let $\{x_n\}$ be a sequence for which $x_n \neq 0$ for all $n \in \mathbf{Z}^+$ and $\{x_n\} \to 0$. Then we can find a number $N \in \mathbf{Z}^+$ such that for all $n \in \mathbf{Z}_{\geq N}(|x_n| < \dfrac{\pi}{2})$. By (9.39)

$$n \in \mathbf{Z}_{\geq N} \implies \cos(x_n) \leq \frac{\sin(x_n)}{x_n} \leq 1.$$

By lemma 9.34, we know that $\{\cos(x_n)\} \to 1$, so by the squeezing rule $\left\{\dfrac{\sin(x_n)}{x_n}\right\} \to 1$. |||

9.40 Example (Calculation of π.) Since $\left\{\dfrac{\pi}{n}\right\} \to 0$, it follows from (9.38) that

$$\lim\left\{\frac{\sin\left(\frac{\pi}{n}\right)}{\frac{\pi}{n}}\right\} = 1$$

and hence that

$$\lim\left\{n\sin\left(\frac{\pi}{n}\right)\right\} = \pi.$$

This result can be used to find a good approximation to π. By the half-angle formula, we have

$$\sin^2\left(\frac{t}{2}\right) = \frac{1-\cos t}{2} = \frac{1}{2}\Big(1 - \sqrt{1-\sin^2 t}\Big)$$

for $0 \leq t \leq \dfrac{\pi}{2}$. Here I have used the fact that $\cos t \geq 0$ for $0 \leq t \leq \dfrac{\pi}{2}$. Also $\sin(\dfrac{\pi}{2}) = 1$ so

$$\sin^2(\frac{\pi}{4}) \;=\; \frac{1}{2}\Big(1 - \sqrt{1-\sin^2\frac{\pi}{2}}\Big) = \frac{1}{2}\big(1-\sqrt{0}\big) = \frac{1}{2}.$$

$$\sin^2(\frac{\pi}{8}) = \frac{1}{2}\Big(1 - \sqrt{1 - \sin^2\frac{\pi}{4}}\Big) = \frac{1}{2}\Big(1 - \sqrt{1 - \frac{1}{2}}\Big) = \frac{1}{2}\Big(1 - \sqrt{\frac{1}{2}}\Big).$$

By repeated applications of this process I can find $\sin^2\left(\frac{\pi}{2^n}\right)$ for arbitrary n, and then find

$$2^n \sin\left(\frac{\pi}{2^n}\right)$$

which will be a good approximation to π.

I wrote a set of Maple routines to do the calculations above. The procedure `sinsq(n)` calculates $\sin^2\left(\frac{\pi}{2^n}\right)$ and the procedure `mypi(m)` calculates $2^m \sin\left(\frac{\pi}{2^m}\right)$. The "fi" (which is "if" spelled backwards) is Maple's way of ending an "if" statement. "Digits := 20" indicates that all calculations are done to 20 decimal digits accuracy. The command "evalf(Pi)" requests the decimal approximation to π to be printed.

```
> sinsq :=
>    n-> if n=1 then 1;
>        else .5*(1-sqrt(1 - sinsq(n-1)));
>        fi;
```

sinsq := proc(n) options operator,arrow; if n = 1 then 1 else .5 -.5*sqrt(1-sinsq(n-1)) fi end

```
> mypi := m -> 2^m*sqrt(sinsq(m));
```

$$mypi := m \to 2^m \operatorname{sqrt}(\operatorname{sinsq}(m))$$

```
> Digits := 20;
```

$$Digits := 20$$

```
> mypi(4);
```

3.1214451522580522853

```
> mypi(8);
```

3.1415138011443010542

```
> mypi(12);
```

3.1415923455701030907

```
> mypi(16);
```

3.1415926523835057093

```
> mypi(20);
                    3.1415926533473327481
> mypi(24);
                    3.1415922701132732445
> mypi(28);
                    3.1414977446171452114
> mypi(32);
                    3.1267833885746006944
> mypi(36);
                              0
> mypi(40);
                              0
> evalf(Pi);
                    3.1415926535897932385
```

9.41 Exercise. Examine the output of the program above. It appears that $\pi = 0$. This certainly is not right. What can I conclude about π from my computer program?

9.42 Exercise. Show that the number $n \sin\left(\frac{\pi}{n}\right)$ is the area of a regular $2n$-gon inscribed in the unit circle. Make any reasonable geometric assumptions, but explain your ideas clearly.

9.3 Integrals of the Trigonometric Functions

9.43 Theorem (Integral of cos**)** *Let $[a, b]$ be an interval in* **R**. *Then the cosine function is integrable on $[a, b]$, and*

$$\int_a^b \cos = \sin(b) - \sin(a).$$

Proof: Let $[a, b]$ be any interval in **R**. Then cos is piecewise monotonic on $[a, b]$ and hence is integrable. Let $P_n = \{x_0, x_1, \cdots, x_n\}$ be the regular partition of $[a, b]$ into n equal subintervals, and let

$$S_n = \left\{\frac{x_0 + x_1}{2}, \frac{x_1 + x_2}{2}, \cdots, \frac{x_{n-1} + x_n}{2}\right\}$$

be the sample for P_n consisting of the midpoints of the intervals of P_n.

Let $\Delta_n = \dfrac{b-a}{n}$ so that $x_i - x_{i-1} = \Delta_n$ and $\dfrac{x_{i-1}+x_i}{2} = x_{i-1} + \dfrac{\Delta_n}{2}$ for $1 \leq i \leq n$. Then

$$\begin{aligned}\sum(\cos, P_n, S_n) &= \sum_{i=1}^{n} \cos\Big(x_{i-1} + \frac{\Delta_n}{2}\Big) \cdot \Delta_n \\ &= \Delta_n \sum_{i=1}^{n} \cos\Big(x_{i-1} + \frac{\Delta_n}{2}\Big).\end{aligned}$$

Multiply both sides of this equation by $\sin\Big(\dfrac{\Delta_n}{2}\Big)$ and use the identity

$$\sin(t)\cos(s) = \frac{1}{2}[\sin(s+t) - \sin(s-t)]$$

to get

$$\begin{aligned}\sin\Big(\frac{\Delta_n}{2}\Big)\sum(\cos, P_n, S_n) &= \Delta_n \sum_{i=1}^{n} \sin\Big(\frac{\Delta_n}{2}\Big)\cos\Big(x_{i-1} + \frac{\Delta_n}{2}\Big) \\ &= \Delta_n \sum_{i=1}^{n} \frac{1}{2}[\sin(x_{i-1} + \Delta_n) - \sin(x_{i-1})] \\ &= \frac{\Delta_n}{2} \sum_{i=1}^{n} \sin(x_i) - \sin(x_{i-1}) \\ &= \frac{\Delta_n}{2}\Big[\Big(\sin(x_n) - \sin(x_{n-1})\Big) + \Big(\sin(x_{n-1}) - \sin(x_{n-2})\Big) \\ &\quad + \cdots + \Big(\sin(x_1) - \sin(x_0)\Big)\Big] \\ &= \frac{\Delta_n}{2}[\sin(x_n) - \sin(x_0)] \\ &= \frac{\Delta_n}{2}\Big(\sin(b) - \sin(a)\Big).\end{aligned}$$

Thus

$$\sum(\cos, P_n, S_n) = \frac{\Big(\frac{\Delta_n}{2}\Big)}{\sin\Big(\frac{\Delta_n}{2}\Big)}\Big(\sin(b) - \sin(a)\Big).$$

(By taking n large enough we can guarantee that $\dfrac{\Delta_n}{2} < \pi$, and then $\sin\Big(\dfrac{\Delta_n}{2}\Big) \neq 0$, so we haven't divided by 0.) Thus by theorem 9.37

$$\begin{aligned}
\int_a^b \cos &= \lim\{\sum(\cos, P_n, S_n)\} \\
&= \lim\left\{\Big(\sin(b)-\sin(a)\Big)\Big(\frac{\frac{\Delta_n}{2}}{\sin\left(\frac{\Delta_n}{2}\right)}\Big)\right\}. \\
&= \Big(\sin(b)-\sin(a)\Big)\cdot\lim\left\{\Big(\frac{\frac{\Delta_n}{2}}{\sin\left(\frac{\Delta_n}{2}\right)}\Big)\right\} \\
&= \Big(\sin(b)-\sin(a)\Big)\cdot 1 = \sin(b)-\sin(a). \ \|
\end{aligned}$$

9.44 Exercise. Let $[a, b]$ be an interval in $\mathbf{R}$. Show that

$$\int_a^b \sin = \cos(a) - \cos(b). \tag{9.45}$$

The proof is similar to the proof of (9.43). The magic factor $\sin\left(\frac{\Delta_n}{2}\right)$ is the same as in that proof.

9.46 Notation ($\int_b^a f$.) If f is integrable on the interval $[a, b]$, we define

$$\int_b^a f = -\int_a^b f \text{ or } \int_b^a f(t)dt = -\int_a^b f(t)dt.$$

This is a natural generalization of the convention for $A_b^a f$ in definition 5.67.

9.47 Theorem (Integrals of sin and cos.) *Let a and b be any real numbers. Then*

$$\int_a^b \cos = \sin(b) - \sin(a).$$

and

$$\int_a^b \sin = \cos(a) - \cos(b).$$

Proof: We will prove the first formula. The proof of the second is similar. If $a \leq b$ then the conclusion follows from theorem 9.43.

If $b < a$ then

$$\int_a^b \cos = -\int_b^a \cos = -[\sin(a) - \sin(b)] = \sin(b) - \sin(a),$$

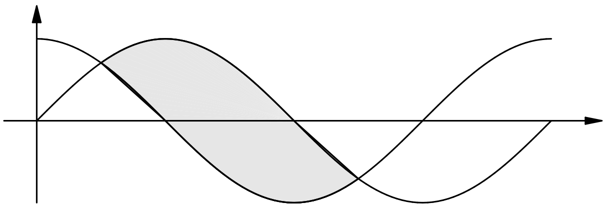
so the conclusion follows in all cases. ‖

9.48 Exercise. Find the area of the set

$$S_0^{\pi}(\sin) = \{(x,y)\colon 0 \le x \le \pi \text{ and } 0 \le y \le \sin x\}.$$

Draw a picture of $S_0^{\pi}(\sin)$.

9.49 Exercise. Find the area of the shaded figure, which is bounded by the graphs of the sine and cosine functions.



9.50 Example. By the change of scale theorem we have for $a < b$ and $c > 0$.

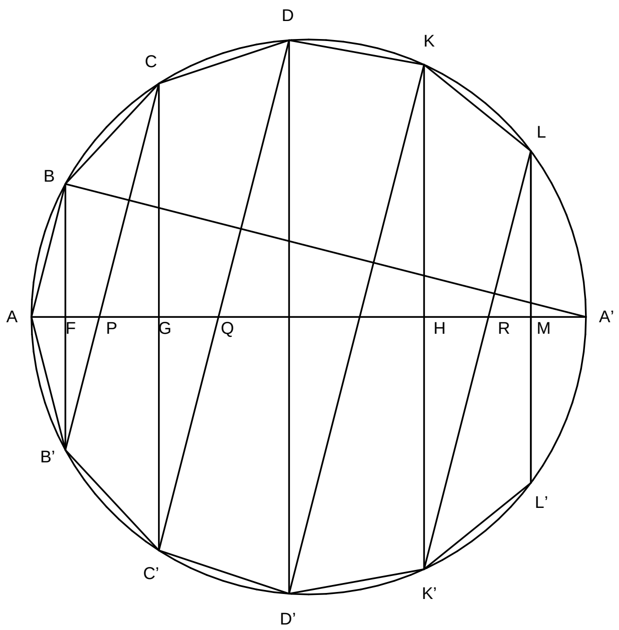
$$\begin{aligned}\int_a^b \sin(cx)dx &= \frac{1}{c}\int_{ca}^{cb} \sin x\, dx \\ &= \frac{-\cos(cb)+\cos(ca)}{c}\end{aligned}$$

$$\begin{aligned}\int_a^b \cos(cx)dx &= \frac{1}{c}\int_{ca}^{cb} \cos x\, dx \\ &= \frac{\sin(cb)-\sin(ca)}{c}\end{aligned}$$

9.51 Entertainment (Archimedes sine integral) In *On the Sphere and Cylinder 1.*, Archimedes states the following proposition: (see figure on next page)

Statement A:

> If a polygon be inscribed in a segment of a circle LAL' so that all its sides excluding the base are equal and their number even, as $LK\ldots A\ldots K'L'$, A being the middle point of segment, and if the lines BB', CC',... parallel to the base LL' and joining pairs of angular points be drawn, then
>
> $$(BB' + CC' + \ldots + LM) : AM = A'B : BA,$$
>
> where M is the middle point of LL' and AA' is the diameter through M.[2, page 29]



We will now show that this result can be reformulated in modern notation as follows.
Statement B: Let ϕ be a number in $[0, \pi]$, and let n be a positive integer. Then there exists a partitition-sample sequence $(\{P_n\}, \{S_n\})$ for $[0, \phi]$, such that

$$\sum(\sin, P_n, S_n) = (1 - \cos(\phi))\frac{\phi}{2n+1}\ \frac{\cos(\frac{\phi}{2n})}{\sin(\frac{\phi}{2n})}. \tag{9.52}$$

In exercise (9.56) you are asked to show that (9.52) implies that

$$\int_0^\phi \sin = 1 - \cos(\phi).$$

Proof that statement A implies statement B: Assume that statement A is true. Take the circle to have radius equal to 1, and let

$$\begin{aligned} \phi &= \text{length of arc}(AL) \\ \frac{\phi}{n} &= \text{length of arc}(AB). \end{aligned}$$

Then

$$BB' + CC' + \ldots + LM = 2\sin(\frac{\phi}{n}) + 2\sin(\frac{2\phi}{n}) + \cdots + 2\sin(\frac{(n-1)\phi}{n}) + \sin(\phi),$$

and

$$AM = 1 - \cos(\phi).$$

Let

$$P_n = \{0, \frac{2\phi}{2n+1}, \frac{4\phi}{2n+1}, \cdots, \frac{2n\phi}{2n+1}, \phi\},$$

and

$$S_n = \{0, \frac{\phi}{n}, \frac{2\phi}{n}, \cdots, \frac{n\phi}{n}\}.$$

Then P_n is a partition of $[0, \phi]$ with mesh equal to $\frac{2\phi}{2n+1}$, and S_n is a sample for P_n, so $(\{P_n\}, \{S_n\})$ is a partition-sample sequence for $[0, \phi]$, and we have

$$\sum(\sin, P_n, S_n) = \frac{2\phi}{2n+1}\Big(\sin(\frac{\phi}{n}) + \sin(\frac{2\phi}{n}) + \cdots + \sin(\frac{(n-1)\phi}{n}) + \frac{1}{2}\sin(\phi)\Big).$$

By Archimedes' formula, we conclude that

$$\sum(\sin, P_n, S_n) = (1 - \cos(\phi))\frac{\phi}{2n+1} \cdot \frac{A'B}{BA}. \tag{9.53}$$

We have

$$\begin{aligned} \text{length arc}(BA) &= \frac{\phi}{n}, \\ \text{length arc}(BA') &= \pi - \frac{\phi}{n}. \end{aligned}$$

By the formula for the length of a chord (9.30) we have

$$\frac{A'B}{BA} = \frac{\text{chord}(AB')}{\text{chord}(BA)} = \frac{2\sin(\frac{\text{arc}(AB')}{2})}{2\sin(\frac{\text{arc}(BA)}{2})} = \frac{\sin(\frac{(\pi - \frac{\phi}{n})}{2}))}{\sin(\frac{(\frac{\phi}{n})}{2})} = \frac{\cos(\frac{\phi}{2n})}{\sin(\frac{\phi}{2n})} \tag{9.54}$$

Equation (9.52) follows from (9.53) and (9.54).

Prove statement A above. Note that (see the figure from statement A)

$$AM = AF + FP + PG + GQ + \cdots + HR + RM, \tag{9.55}$$

and each summand on the right side of (9.55) is a side of a right triangle similar to triangle $A'BA$.

9.56 Exercise. Assuming equation (9.52), show that

$$\int_0^\phi \sin = 1 - \cos(\phi).$$

9.4 Indefinite Integrals

9.57 Theorem. *Let a, b, c be real numbers. If f is a function that is integrable on each interval with endpoints in $\{a, b, c\}$ then*

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Proof: The case where $a \leq b \leq c$ is proved in theorem 8.18. The rest of the proof is exactly like the proof of exercise 5.69. |||

9.58 Exercise. Prove theorem 9.57.

We have proved the following formulas:

$$\int_a^b x^r dx = \frac{b^{r+1} - a^{r+1}}{r+1} \text{ for } 0 < a < b \quad r \in \mathbf{Q} \setminus \{-1\}, \tag{9.59}$$

$$\int_a^b \frac{1}{t} dt = \ln(b) - \ln(a) \text{ for } 0 < a < b,$$

$$\int_a^b \sin(ct)dt = \frac{-\cos(cb) + \cos(ca)}{c} \text{ for } a < b, \text{ and } c > 0,$$

$$\int_a^b \cos(ct)dt = \frac{\sin(cb) - \sin(ca)}{c} \text{ for } a < b, \text{ and } c > 0. \tag{9.60}$$

In each case we have a formula of the form

$$\int_a^b f(t)dt = F(b) - F(a).$$

This is a general sort of situation, as is shown by the following theorem.

9.61 Theorem (Existence of indefinite integrals.) *Let J be an interval in $\mathbf{R}$, and let $f: J \to \mathbf{R}$ be a function such that f is integrable on every subinterval $[p, q]$ of J. Then there is a function $F: J \to \mathbf{R}$ such that for all $a, b \in J$*

$$\int_a^b f(t)dt = F(b) - F(a).$$

Proof: Choose a point $c \in J$ and define

$$F(x) = \int_c^x f(t)dt \text{ for all } x \in J.$$

Then for any points a, b in J we have

$$F(b) - F(a) = \int_c^b f(t)dt - \int_c^a f(t)dt = \int_a^b f(t)dt.$$

We've used the fact that

$$\int_c^b f(t)dt = \int_c^a f(t)dt + \int_a^b f(t)dt \text{ for all } a, b, c \in J. \ \|$$

9.62 Definition (Indefinite integral.) Let f be a function that is integrable on every subinterval of an interval J. An *indefinite integral for f on J* is any function $F\colon J \to \mathbf{R}$ such that $\int_a^b f(t)dt = F(b) - F(a)$ for all $a, b \in J$.

A function that has an indefinite integral always has infinitely many indefinite integrals, since if F is an indefinite integral for f then so is $F + c$ for any number c:

$$(F + c)(b) - (F + c)(a) = (F(b) + c) - (F(a) + c) = F(b) - F(a).$$

The following notation is used for indefinite integrals. One writes $\int f(t)dt$ to denote an indefinite integral for f. The t here is a dummy variable and can be replaced by any available symbol. Thus, based on formulas (9.59) - (9.60), we write

$$\begin{aligned}
\int x^r dx &= \frac{x^{r+1}}{r+1} \text{ if } r \in \mathbf{Q} \setminus \{-1\} \\
\int \frac{1}{t} dt &= \ln(t) \\
\int \sin(ct)dt &= -\frac{\cos(ct)}{c} \text{ if } c > 0 \\
\int \cos(ct)dt &= \frac{\sin(ct)}{c} \text{ if } c > 0.
\end{aligned}$$

We might also write

$$\int x^r dr = \frac{x^{r+1}}{r+1} + 3.$$

Some books always include an arbitrary constant with indefinite integrals, e.g.,

$$\int x^r dr = \frac{x^{r+1}}{r+1} + C \text{ if } r \in \mathbf{Q} \setminus \{-1\}.$$

The notation for indefinite integrals is treacherous. If you see the two equations

$$\int x^3 dx = \frac{1}{4}x^4$$

and

$$\int x^3 dx = \frac{1}{4}(x^4+1),$$

then you want to conclude

$$\frac{1}{4}x^4 = \frac{1}{4}(x^4+1), \tag{9.63}$$

which is wrong. It would be more logical to let the symbol $\int f(x)dx$ denote the set of *all* indefinite integrals for f. If you see the statements

$$\frac{1}{4}x^4 \in \int x^3 dx$$

and

$$\frac{1}{4}(x^4+1) \in \int x^3 dx,$$

you are not tempted to make the conclusion in (9.63).

9.64 Theorem (Sum theorem for indefinite integrals) *Let f and g be functions each of which is integrable on every subinterval of an interval J, and let $c, k \in \mathbf{R}$. Then*

$$\int \Big(cf(x) + kg(x)\Big)dx = c\int f(x)dx + k\int g(x)dx. \tag{9.65}$$

Proof: The statement (9.65) means that if F is an indefinite integral for f and G is an indefinite integral for G, then $cF + kG$ is an indefinite integral for $cf + kg$.

Let F be an indefinite integral for f and let G be an indefinite integral for g. Then for all $a, b \in J$

$$\begin{aligned}
\int_a^b \Big(cf(x) + kg(x)\Big)dx &= \int_a^b cf(x)dx + \int_a^b kg(x)dx \\
&= c\int_a^b f(x)dx + k\int_a^b g(x)dx \\
&= c\Big(F(b) - F(a)\Big) + k\Big(G(b) - G(a)\Big) \\
&= \Big(cF(b) + kG(b)\Big) - \Big(cF(a) + kG(a)\Big) \\
&= (cF + kG)(b) - (cF + kG)(a).
\end{aligned}$$

It follows that $cF + kG$ is an indefinite integral for $cf + kg$. |||

9.66 Notation ($F(t)\ |_a^b$.) If F is a function defined on an interval J, and if a, b are points in J we write $F(t)\ |_a^b$ for $F(b) - F(a)$. The t here is a dummy variable, and sometimes the notation is ambiguous, e.g. $x^2 - t^2\ |_0^1$. In such cases we may write $F(t)\ |_{t=a}^{t=b}$. Thus

$$(x^2 - t^2)\ |_{x=0}^{x=1} = (1 - t^2) - (0 - t^2) = 1$$

while

$$(x^2 - t^2)\ |_{t=0}^{t=1} = (x^2 - 1) - (x^2 - 0) = -1.$$

Sometimes we write $F\ |_a^b$ instead of $F(t)\ |_a^b$.

9.67 Example. It follows from our notation that if F is an indefinite integral for f on an interval J then

$$\int_a^b f(t)dt = F(t)\ |_a^b$$

and this notation is used as follows:

$$\begin{aligned}
\int_a^b 3x^2 dx &= x^3\Big|_a^b = b^3 - a^3. \\
\int_0^\pi \cos(x)dx &= \sin(x)\Big|_0^\pi = 0 - 0 = 0. \\
\int_0^\pi \sin(3x)dx &= \frac{-\cos 3x}{3}\Big|_0^\pi = \frac{-\cos(3\pi)}{3} + \frac{\cos(0)}{3} = \frac{2}{3}. \\
\int_0^2 (4x^2 + 3x + 1)dx &= 4\Big(\frac{x^3}{3}\Big) + 3\Big(\frac{x^2}{2}\Big) + x\Big|_0^2 \\
&= 4 \cdot \frac{8}{3} + 3 \cdot \frac{4}{2} + 2 = \frac{56}{3}.
\end{aligned}$$

In the last example I have implicitly used

$$\int (4x^2 + 3x + 1)dx = 4\int x^2 dx + 3\int x\ dx + \int 1\ dx.$$

9.68 Example. By using the trigonometric identities from theorem 9.21 we can calculate integrals of the form $\int_a^b \sin^n(cx)\cos^m(kx)dx$ where m, n are non-negative integers and $c, k \in \mathbf{R}$. We will find

$$\int_0^{\frac{\pi}{2}} \sin^3(x) \cdot \cos(3x)dx.$$

We have

$$\sin^2(x) = \frac{1-\cos(2x)}{2},$$

so

$$\begin{aligned}\sin^3(x) &= \sin^2(x)\sin(x) = \frac{1}{2}\sin(x) - \frac{1}{2}\cos(2x)\sin(x)\\ &= \frac{1}{2}\sin(x) - \frac{1}{2}\cdot\frac{1}{2}\Big(\sin(3x)-\sin(x)\Big)\\ &= \frac{3}{4}\sin(x) - \frac{1}{4}\sin(3x).\end{aligned}$$

Thus

$$\begin{aligned}\sin^3(x)\cdot\cos(3x) &= \frac{3}{4}\cos(3x)\sin(x) - \frac{1}{4}\cos(3x)\sin(3x)\\ &= \frac{3}{8}[\sin(4x)-\sin(2x)] - \frac{1}{8}\sin(6x).\end{aligned}$$

Hence

$$\int_0^{\pi/2} \sin^3(x)\cdot\cos(3x)\,dx$$

$$\begin{aligned}&= \frac{3}{8}\frac{(-\cos(4x))}{4}\bigg|_0^{\frac{\pi}{2}} - \frac{3}{8}\frac{(-\cos(2x))}{2}\bigg|_0^{\frac{\pi}{2}} - \frac{1}{8}\frac{(-\cos(6x))}{6}\bigg|_0^{\frac{\pi}{2}}\\ &= \frac{3}{32}\Big(-\cos(2\pi)+\cos(0)\Big) + \frac{3}{16}\Big(\cos(\pi)-\cos(0)\Big) + \frac{1}{48}\Big(\cos(3\pi)-\cos(0)\Big)\\ &= \frac{3}{16}(-1-1) + \frac{1}{48}(-1-1) = -\frac{3}{8} - \frac{1}{24} = \frac{-10}{24} = -\frac{5}{12}.\end{aligned}$$

The method here is clear, but a lot of writing is involved, and there are many opportunities to make errors. In practice I wouldn't do a calculation of this sort by hand. The Maple command

```
> int((sin(x))^3*cos(3*x),x=0..Pi/2);
```

responds with the value

```
- 5/12
```

9.69 Exercise. Calculate the integrals

$$\int_0^{\frac{\pi}{2}} \sin x\, dx,\ \int_0^{\frac{\pi}{2}} \sin^2 x\, dx \text{ and } \int_0^{\frac{\pi}{2}} \sin^4 x\, dx.$$

Then determine the values of

$$\int_0^{\frac{\pi}{2}} \cos x\, dx,\ \int_0^{\frac{\pi}{2}} \cos^2 x\, dx \text{ and } \int_0^{\frac{\pi}{2}} \cos^4 x\, dx$$

without doing any calculations. (But include an explanation of where your answer comes from.)

9.70 Exercise. Find the values of the following integrals. If the answer is geometrically clear then don't do any calculations, but explain why the answer is geometrically clear.

a) $\displaystyle\int_1^2 \frac{1}{x^3} dx$

b) $\displaystyle\int_{-1}^1 x^{11}(1+x^2)^3\, dx$

c) $\displaystyle\int_0^2 \sqrt{4-x^2}\, dx$

d) $\displaystyle\int_0^{\pi} (x+\sin(2x))dx$

e) $\displaystyle\int_{-1}^1 \frac{1}{x^2} dx$

f) $\displaystyle\int_1^4 \frac{4+x}{x} dx$

g) $\displaystyle\int_0^1 \sqrt{x} dx$

h) $\displaystyle\int_1^2 \frac{4}{x} dx$

i) $\displaystyle\int_0^1 (1-2x)^2 dx$

j) $\displaystyle\int_0^1 (1-2x) dx$

k) $\int_0^{\pi} \sin(7x)\, dx$

l) $\int_0^{\pi} \sin(8x)\, dx$

9.71 Exercise.

$$\begin{aligned} \text{Let } A &= \int_0^{\pi/2} (\sin(4x))^5 dx \\ B &= \int_0^{\pi/2} (\sin(3x))^5 dx \\ C &= \int_0^{\frac{\pi}{2}} (\cos(3x))^5 dx. \end{aligned}$$

Arrange the numbers A, B, C in increasing order. Try to do the problem without making any explicit calculations. By making rough sketches of the graphs you should be able to come up with the answers. Sketch the graphs, and explain how you arrived at your conclusion. No "proof" is needed.

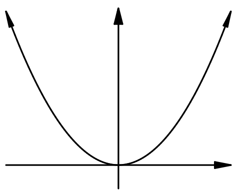
Chapter 10

Definition of the Derivative

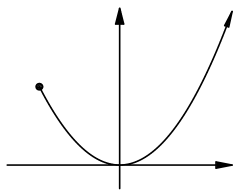
10.1 Velocity and Tangents

10.1 Notation. If $E_1(x, y)$ and $E_2(x, y)$ denote equations or inequalities in x and y, we will use the notation

$$\begin{aligned} \{E_1(x,y)\} &= \{(x,y) :\in \mathbf{R}^2 \colon E_1(x,y)\} \\ \{E_1(x,y); E_2(x,y)\} &= \{(x,y) \in \mathbf{R}^2 \colon E_1(x,y) \text{ and } E_2(x,y)\}. \end{aligned}$$



$\{y = x^2\}$



$\{y = x^2; x \geq 1\}$

In this section we will discuss the idea of *tangent to a curve* and the related idea of *velocity of a moving point.*

You probably have a pretty good intuitive idea of what is meant by the tangent to a curve, and you can see that the straight lines in figure a below represent tangent lines to curves.

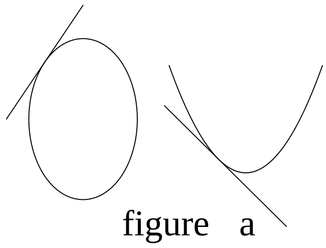
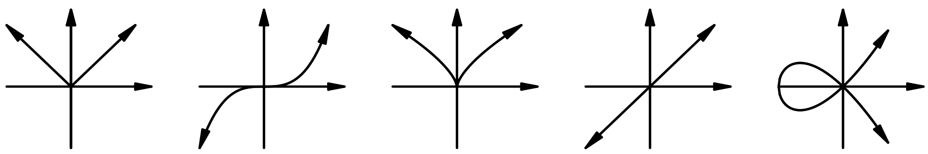


figure a

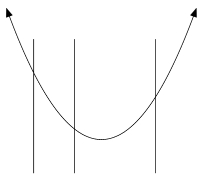
It may not be quite so clear what you would mean by the tangents to the curves in figure b at the point $(0, 0)$.



$\{y = |x|\}$ $\{y = x^3\}$ $\{y^3 = x^2\}$ $\{y = x\}$ $\{y^2 = x^3 + x^2\}$

figure b

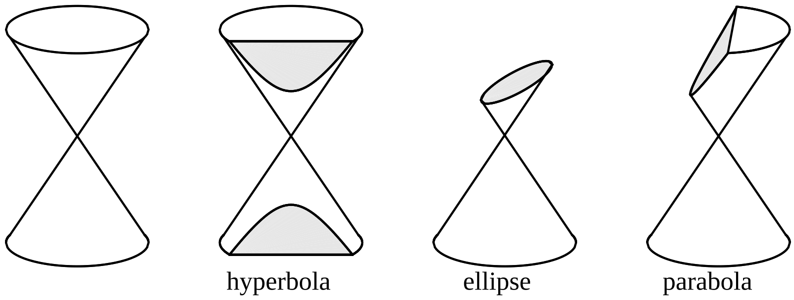
Euclid (fl. c. 300 B.C.) defined a tangent to a circle to be a line which touches the circle in exactly one point. This is a satisfactory definition of tangent to a circle, but it does not generalize to more complicated curves.



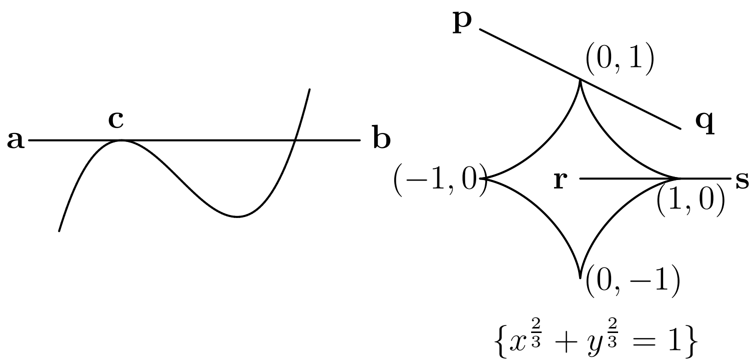
$\{y = x^2\}$

For example, every vertical line intersects the parabola $\{y = x^2\}$ in just one point, but no such line should be considered to be a tangent.

Apollonius (c 260-170 B.C.) defined a tangent to a conic section (i.e., an ellipse or hyperbola or parabola) to be a line that touches the section, but lies outside of the section. Apollonius considered these sections to be obtained by intersecting a cone with a plane, and points inside of the section were points in the cone.



This definition works well for conic sections, but for general curves, we have no notion of what points lie inside or outside a curve.

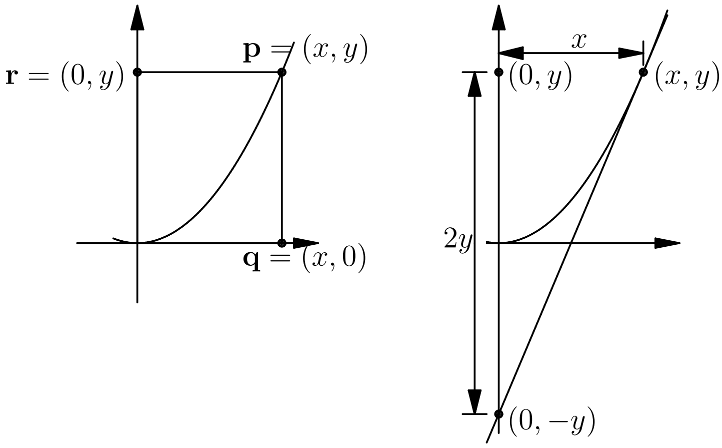


In the figure, the line **ab** ought to be tangent to the curve at **c**, but there is no reasonable sense in which the line lies outside the curve. On the other hand, it may not be clear whether **pq** (which lies outside the curve $\{x^{2/3} + y^{2/3} = 1\}$ is more of a tangent than the line **rs** which does not lie outside of it. Leibniz [33, page 276] said that

> to find a *tangent* means to draw a line that connects two points of the curve at an infinitely small distance, or the continued side of a polygon with an infinite number of angles, which for us takes the place of the *curve*.

From a modern point of view it is hard to make any sense out of this.

Here is a seventeenth century sort of argument for finding a tangent to the parabola whose equation is $y = x^2$.



Imagine a point **p** that is moving along the parabola $y = x^2$, so that at time t, **p** is at (x, y). (Here x and y are functions of t, but in the seventeenth century they were just flowing quantities.) Imagine a point **q** that moves along the x-axis so that **q** always lies under **p** and a point **r** moving along the y-axis so that **r** is always at the same height as **p**. Let $\dot{x}$ denote the velocity of **q** when **p** is at (x, y) and let $\dot{y}$ denote the velocity of **r** when **p** is at (x, y). Let o be a very small moment of time. At time o after **p** is at (x, y), **p** will be at $(x + o\dot{x}, y + o\dot{y})$ (i.e., x will have increased by an amount equal to the product of the time interval o and its velocity $\dot{x}$). Since **p** stays on the curve, we have

$$y + o\dot{y} = (x + o\dot{x})^2$$

or

$$y + o\dot{y} = x^2 + 2xo\dot{x} + o^2\dot{x}^2.$$

Since $y = x^2$, we get

$$o\dot{y} = 2xo\dot{x} + o^2\dot{x}^2 \tag{10.2}$$

or

$$\dot{y} = 2x\dot{x} + o\dot{x}^2 \tag{10.3}$$

Since we are interested in the velocities at the instant that **p** is at (x, y), we take $o = 0$, so

$$\dot{y} = 2x\dot{x}.$$

Hence when p is at (x, y), the vertical part of its velocity (i.e., $\dot{y}$) is $2x$ times the horizontal component of its velocity. Now the velocity should point in the direction of the curve; i.e., in the direction of the tangent, so the direction of the tangent at (x, y) should be in the direction of the diagonal of a box with

$$\text{vertical side } = 2x \times \text{ horizontal side.}$$

The tangent to the parabola at $(x, y) = (x, x^2)$ is the line joining (x, y) to $(0, -y)$, since in the figure the vertical component of the box is

$$2y = 2x^2 = (2x)x;$$

i.e., the vertical component is $2x$ times the horizontal component.

In *The Analyst: A Discourse Addressed to an Infidel Mathematician*[7, page 73], George Berkeley (1685-1753) criticizes the argument above, pointing out that when we divide by o in line (10.3) we must assume o is not zero, and then at the end we set o equal to 0.

> All which seems a most inconsistent way of arguing, and such as would not be allowed of in Divinity.

The technical concept of velocity is not a simple one. The idea of *uniform velocity* causes no problems: to quote Galileo (1564-1642):

> By steady or uniform motion, I mean one in which the distances traversed by the moving particle during any equal intervals of time, are themselves equal[21, page 154].

This definition applies to points moving in a straight line, or points moving on a circle, and it goes back to the Greek scientists. The problem of what is meant by velocity for a non-uniform motion, however, is not at all clear. The Greeks certainly realized that a freely falling body moves faster as it falls, but they had no language to describe the way in which velocity changes. Aristotle (384-322 B.C.) says

> there cannot be motion of motion or becoming of becoming or in general change of change[11, page 168].

It may not be clear what this means, but S. Bochner interprets this as saying that the notion of a second derivative (this is a technical term for the mathematical concept used to describe acceleration which we will discuss later) is a meaningless idea[11, page 167]. Even though we are in constant contact with non-uniformly moving bodies, our intuition about the way they move is not very good. In the *Dialogues Concerning Two New Sciences*, Salviati (representing Galileo) proposes the hypothesis that if a stone falls from rest, then it falls in such a way that "in any equal intervals of time whatever, equal increments of speed are given to it"[21, page 161].

In our language, the hypothesis is that the velocity $v(t)$ at time t satisfies

$$v(t) = kt \text{ for some constant } k.$$

Sagredo objects to this on the grounds that this would mean that the object begins to fall with zero speed "while our senses show us that a heavy falling body suddenly acquires great speed." (I believe Sagredo is right. Try dropping some bodies and observe how they begin to fall.) Salviati replies that this is what he thought at first, and explains how he came to change his mind.

Earlier, in 1604, Galileo had supposed that

$$v(x) = kx \text{ for some constant } k;$$

i.e., that in equal increments of distance the object gains equal increments of speed (which is false), and Descartes made the same error in 1618 [13, page 165]. Casual observation doesn't tell you much about falling stones.

10.4 Entertainment (Falling bodies.) Try to devise an experiment to support (or refute) Galileo's hypothesis that $v(t) = kt$, using materials available to Galileo; e.g., no stop watch. Galileo describes his experiments in [21, pages 160-180], and it makes very good reading.

10.2 Limits of Functions

Our definition of tangent to a curve is going to be based on the idea of *limit*. The word *limit* was used in mathematics long before the definition we will give was thought of. One finds statements like "The limit of a regular polygon when the number of sides becomes infinite, is a circle." Early definitions of limit often involved the ideas of time or motion. Our definition will be purely mathematical.

10.5 Definition (Interior points and approachable points.) Let S be a subset of $\mathbf{R}$. A point $x \in S$ is an *interior point* of S if there is some positive number ϵ such that the interval $(x - \epsilon, x + \epsilon)$ is a subset of S. A point $x \in R$

is an *approachable point from* S if there is some positive number ϵ such that either $(x-\epsilon, x) \subset S$ or $(x, x+\epsilon) \subset S$. (Without loss of generality we could replace "ϵ" in this definition by $\dfrac{1}{N}$ for some $N \in \mathbf{Z}^+$.)

Note that interior points of S must belong to S. Approachable points of S need not belong to S. Any interior point of S is approachable from S.

10.6 Example. If S is the open interval $(0,1)$ then every point of S is an interior point of S. The points that are approachable from S are the points in the closed interval $[0,1]$.

If T is the closed interval $[0,1]$ then the points that are approachable from T are exactly the points in T, and the interior points of T are the points in the open interval $(0,1)$.

10.7 Definition (Limit of a function.) Let f be a real valued function such that $\text{dom}(f) \subset \mathbf{R}$. Let $a \in \mathbf{R}$ and let $L \in \mathbf{R}$. We say

$$\lim_{x \to a} f(x) = L \tag{10.8}$$

if

1) a is approachable from $\text{dom}(f)$, and

2) For every sequence $\{x_n\}$ in $\text{dom}(f) \setminus \{a\}$

$$\{x_n\} \to a \quad \Longrightarrow \quad \{f(x_n)\} \to L.$$

Note that the value of $f(a)$ (if it exists) has no influence on the meaning of $\lim_{x \to a} f(x) = L$. Also the "x" in (10.8) is a dummy variable, and can be replaced by any other symbol that has no assigned meaning.

10.9 Example. For all $a \in \mathbf{R}$ we have

$$\lim_{x \to a} x = a.$$

Also

$$\lim_{x \to a} \cos(x) = \cos(a)$$

and

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

by lemma 9.34. Also

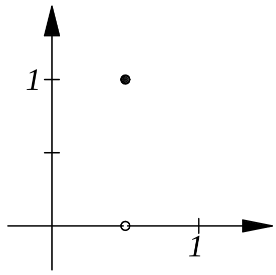
$$\lim_{x\to 0}\frac{\sin x}{x}=1$$

by theorem 9.37.

10.10 Example. $\displaystyle\lim_{x\to 0}\frac{x}{|x|}$ is not defined. Let $x_n=\dfrac{(-1)^n}{n}$. Then $\{x_n\}$ is a sequence in $\mathbf{R}\setminus\{0\}$, and $\{x_n\}\to 0$ and $\dfrac{x_n}{|x_n|}=\dfrac{\frac{(-1)^n}{n}}{\left(\frac{1}{n}\right)}=(-1)^n$. We know there is no number L such that $\{(-1)^n\}\to L$.

10.11 Example. Let f be the spike function

$$f(x)=\begin{cases}0 & \text{if } x\in\mathbf{R}\setminus\{\frac{1}{2}\}\\ 1 & \text{if } x=\frac{1}{2}.\end{cases}$$



Then $\displaystyle\lim_{x\to\frac{1}{2}}f(x)=0$, since if $\{x_n\}$ is a generic sequence in $\text{dom}(f)\setminus\{\frac{1}{2}\}$, then $\{f(x_n)\}$ is the constant sequence $\{0\}$.

10.12 Example. The limit

$$\lim_{x\to 0}(\sqrt{x}+\sqrt{-x})$$

does not exist. If $f(x)=\sqrt{x}+\sqrt{-x}$, then the domain of f consists of the single point 0, and 0 is not approachable from $\text{dom}(f)$. If we did not have condition **1)** in our definition, we would have

$$\lim_{x\to 0}\sqrt{x}+\sqrt{-x}=0 \text{ and } \lim_{x\to 0}\sqrt{x}+\sqrt{-x}=\pi,$$

which would not be a good thing. (If there are no sequences in $\text{dom}(f) \setminus \{a\}$, then

$$\text{for every sequence } \{x_n\} \text{ in } \text{dom}(f) \setminus \{a\}[\text{statement about } \{x_n\}]$$

is true, no matter what [statement about $\{x_n\}$] is.)

In this course we will not care much about functions like $\sqrt{x} + \sqrt{-x}$.

10.13 Example. I will show that

$$\lim_{x \to a} \sqrt{x} = \sqrt{a} \tag{10.14}$$

for all $a \in \mathbf{R}_{\geq 0}$.
Case 1: Suppose $a \in \mathbf{R}^+$. Let $\{x_n\}$ be a generic sequence in $\mathbf{R}^+ \setminus \{a\}$ such that $\{x_n\} \to a$. Then

$$0 \leq |\sqrt{x}_n - \sqrt{a}| = \Big|\frac{\sqrt{x}_n - \sqrt{a}}{1} \cdot \frac{\sqrt{x}_n + \sqrt{a}}{\sqrt{x}_n + \sqrt{a}}\Big| = \frac{|x_n - a|}{\sqrt{x}_n + \sqrt{a}} < \frac{|x_n - a|}{\sqrt{a}}.$$

Now, since $\{x_n\} \to a$, we have

$$\lim \left\{ \frac{|x_n - a|}{\sqrt{a}} \right\} = \frac{1}{\sqrt{a}} \lim\{|x_n - a|\} = 0,$$

so by the squeezing rule $\lim\{\sqrt{x_n} - \sqrt{a}\} = 0$ which is equivalent to

$$\lim\{\sqrt{x_n}\} = \sqrt{a}.$$

This proves (10.14) when $a > 0$.
Case 2: Suppose $a = 0$. The domain of the square root function is $[0, \infty)$, and 0 is approachable from this set.

Let $\{x_n\}$ be a sequence in $\mathbf{R}^+$ such that $\{x_n\} \to 0$. To show that $\{\sqrt{x_n}\} \to 0$, I'll use the definition of limit. Let $\epsilon \in \mathbf{R}^+$. Then $\epsilon^2 \in \mathbf{R}^+$, so by the definition of convergence, there is an $N(\epsilon^2) \in \mathbf{Z}^+$ such that for all $n \in \mathbf{Z}_{\geq N(\epsilon^2)}$ we have $(x_n = |x_n - 0| < \epsilon^2)$. Then for all $n \in \mathbf{Z}_{\geq N(\epsilon^2)}$ we have $(\sqrt{x_n} = |\sqrt{x_n} - 0| < \epsilon)$ and hence $\{\sqrt{x}_n\} \to 0$. ‖

Many of our rules for limits of sequences have immediate corollaries as rules for limits of functions. For example, suppose f, g are real valued functions with

$\text{dom}(f) \subset \mathbf{R}$ and $\text{dom}(g) \subset \mathbf{R}$. Suppose $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Let $\{x_n\}$ be a generic sequence in $\big(\text{dom}(f) \cap \text{dom}(g)\big) \setminus \{a\}$ such that $\{x_n\} \to a$. Then $\{x_n\}$ is a sequence in $\text{dom}(f) \setminus \{a\}$ and $\{x_n\} \to a$, so

$$\{f(x_n)\} \to L.$$

Also $\{x_n\}$ is a sequence in $\text{dom}(g) \setminus \{a\}$ and $\{x_n\} \to a$ so

$$\{g(x_n)\} \to M.$$

By the sum and product rules for sequences, for any $c \in \mathbf{R}$

$$\begin{aligned} \{(f \pm g)(x_n)\} &= \{f(x_n) \pm g(x_n)\} \to L \pm M, \\ \{(fg)(x_n)\} &= \{f(x_n)g(x_n)\} \to LM, \end{aligned}$$

and

$$\{(cf)(x_n)\} = \{c \cdot f(x_n)\} \to cL,$$

and thus we've proved that

$$\begin{aligned} \lim_{x\to a}(f \pm g)(x) &= L \pm M = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x) \\ \lim_{x\to a}(fg)(x) &= LM = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) \end{aligned}$$

and

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

Moreover if $a \in \text{dom}(\frac{f}{g})$ (so that $g(a) \neq 0$), and if $x_n \in \text{dom}(\frac{f}{g})$ for all x_n (so that $g(x_n) \neq 0$ for all n), it follows from the quotient rule for sequences that

$$\Big\{(\frac{f}{g})(x_n)\Big\} = \Big\{\frac{f(x_n)}{g(x_n)}\Big\} \to \frac{L}{M},$$

so that

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} \text{ if } \lim_{x\to a} g(x) \neq 0.$$

Actually all of the results just claimed are not quite true as stated. For we have

$$\lim_{x\to 0} \sqrt{x} = 0$$

and

$$\lim_{x\to 0} \sqrt{-x} = 0$$

but

$$\lim_{x\to 0} \sqrt{x} + \sqrt{-x} \text{ does not exist!}$$

The correct theorem is:

10.15 Theorem (Sum, product, quotient rules for limits.) *Let f, g be real valued functions with* $\text{dom}(f) \subset \mathbf{R}$ *and* $\text{dom}(g) \subset \mathbf{R}$, *and let* $c \in \mathbf{R}$. *Suppose* $\lim_{x\to a} f(x)$, *and* $\lim_{x\to a} g(x)$ *both exist. Then* if a is approachable from $\text{dom}(f) \cap \text{dom}(g)$ *we have*

$$\begin{aligned}
\lim_{x\to a}(f \pm g)(x) &= \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x) \\
\lim_{x\to a}(fg)(x) &= \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) \\
\lim_{x\to a}(cf)(x) &= c \cdot \lim_{x\to a} f(x).
\end{aligned}$$

If in addition $\lim_{x\to a} g(x) \neq 0$ *then*

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

Proof: Most of the theorem follows from the remarks made above. We will assume the remaining parts.

10.16 Theorem (Inequality rule for limits of functions.) *Let f and g be real functions with* $\text{dom}(f) \subset \mathbf{R}$ *and* $\text{dom}(g) \subset \mathbf{R}$. *Suppose that*

- **i** $\lim_{x\to a} f(x)$ *and* $\lim_{x\to a} g(x)$ *both exist.*
- **ii** *a is approachable from* $\text{dom}(f) \cap \text{dom}(g)$.
- **iii** *There is a positive number ϵ such that*

$$f(x) \leq g(x) \textit{ for all } x \textit{ in } \text{dom}(f) \cap \text{dom}(g) \cap (a - \epsilon, a + \epsilon).$$

Then $\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x)$.

Proof: Let $\{x_n\}$ be a sequence in $\Big(\mathrm{dom}(f)\cap\mathrm{dom}(g)\cap(a-\epsilon,a+\epsilon)\Big)\setminus\{a\}$ such that $\{x_n\}\to a$. Then $\{x_n\}$ is a sequence in $\mathrm{dom}(f)\setminus\{a\}$ that converges to a, so by the definition of limit of a function,

$$\lim\{f(x_n)\} = \lim_{x\to a} f(x).$$

Similiarly

$$\lim\{g(x_n)\} = \lim_{x\to a} g(x).$$

Also $f(x_n)\leq g(x_n)$ for all n, so it follows from the inequality rule for limits of sequences that $\lim\{f(x_n)\}\leq\lim\{g(x_n)\}$, i.e. $\lim\limits_{x\to a} f(x)\leq \lim\limits_{x\to a} g(x)$. ⦀.

10.17 Theorem (Squeezing rule for limits of functions.) *Let f, g and h be real functions with* $\mathrm{dom}(f)\subset\mathbf{R}$, $\mathrm{dom}(g)\subset\mathbf{R}$, *and* $\mathrm{dom}(h)\subset\mathbf{R}$. *Suppose that*

i $\lim\limits_{x\to a} f(x)$ *and* $\lim\limits_{x\to a} h(x)$ *both exist and are equal.*

ii *a is approachable from* $\mathrm{dom}(f)\cap\mathrm{dom}(g)\cap\mathrm{dom}(h)$.

iii *There is a positive number ϵ such that $f(x)\leq g(x)\leq h(x)$ for all x in* $\mathrm{dom}(f)\cap\mathrm{dom}(g)\cap\mathrm{dom}(h)\cap(a-\epsilon,a+\epsilon)$.

Then $\lim\limits_{x\to a} f(x)=\lim\limits_{x\to a} g(x)=\lim\limits_{x\to a} h(x)$.

Proof: The proof is almost identical to the proof of theorem 10.16.

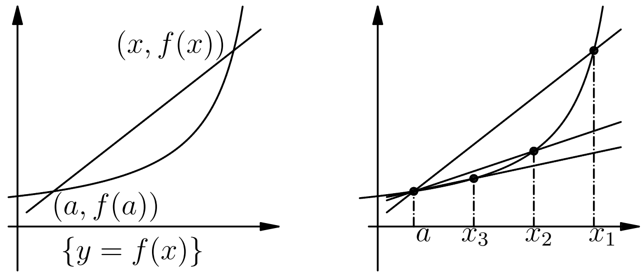
10.3 Definition of the Derivative.

Our definition of tangent to a curve will be based on the following definition:

10.18 Definition (Derivative.) Let f be a real valued function such that $\mathrm{dom}(f)\subset\mathbf{R}$. Let $a\in\mathrm{dom}(f)$. We say that f is *differentiable at a* if a is an interior point of $\mathrm{dom}(f)$ and the limit

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a} \tag{10.19}$$

exists. In this case we denote the limit in (10.19) by $f'(a)$, and we call $f'(a)$ the *derivative of f at a*.



The quantity $\dfrac{f(x)-f(a)}{x-a}$ represents the slope of the line joining the points $\big(a, f(a)\big)$ and $\big(x, f(x)\big)$ on the graph of f. If x and a are different points in $\text{dom}(f)$ then this quotient will be defined. If we choose a sequence of points $\{x_n\}$ converging to a, and if the slopes $\left\{\dfrac{f(x_n)-f(a)}{x_n-a}\right\}$ converge to a number m which is independent of the sequence $\{x_n\}$, then it is reasonable to call m (i.e., $f'(a)$) the *slope of the tangent line to the graph of* f *at* $\big(a, f(a)\big)$.

10.20 Definition (Tangent to the graph of a function.) Let f be a real valued function with $\text{dom}(f) \subset \mathbf{R}$, and let $a \in \text{dom}(f)$. If f is differentiable at a then we define the *slope of the tangent to* $\text{graph}(f)$ at the point $\big(a, f(a)\big)$ to be the number $f'(a)$, and we define the *tangent to* $\text{graph}(f)$ at $\big(a, f(a)\big)$ to be the line that passes through $\big(a, f(a)\big)$ with slope $f'(a)$.

Remark: This definition will need to be generalized later to apply to curves that are not graphs of functions. Also this definition does not allow vertical lines to be tangents, whereas on geometrical grounds, vertical tangents are quite reasonable.

10.21 Example. We will calculate the tangent to $\{y = x^3\}$ at a generic point (a, a^3).

Let $f(x) = x^3$. Then for all $a \in \mathbf{R}$,

$$\begin{aligned}\lim_{x\to a}\frac{f(x)-f(a)}{x-a} &= \lim_{x\to a}\frac{x^3-a^3}{x-a} = \lim_{x\to a}\frac{(x-a)(x^2+ax+a^2)}{(x-a)}\\ &= \lim_{x\to a}(x^2+ax+a^2) = a^2+a^2+a^2 = 3a^2.\end{aligned}$$

Hence the tangent line to graph(f) at (a, a^3) is the line through (a, a^3) with slope $3a^2$, and the equation of the tangent line is

$$y - a^3 = 3a^2(x - a)$$

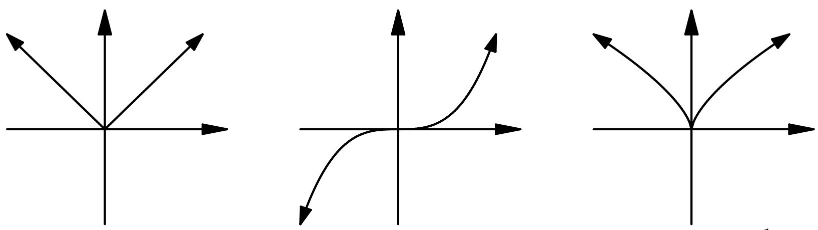
or

$$y = a^3 + 3a^2x - 3a^3 = 3a^2x - 2a^3$$

or

$$y = a^2(3x - 2a).$$

10.22 Example. We will now consider some of the examples on page 220.



$\{y = |x|\}$
no tangent at $(0, 0)$

$\{y = x^3\}$
tangent crosses
curve at $(0, 0)$

$\{y = (x^2)^{\frac{1}{3}}\}$
no tangent at $(0, 0)$

If $f(x) = |x|$ then

$$\frac{f(x) - f(0)}{x - 0} = \frac{|x|}{x}.$$

We saw in example 10.10 that $\lim\limits_{x \to 0} \frac{|x|}{x}$ does not exist. Hence, the graph of f at $(0, 0)$ has no tangent.

If $g(x) = x^3$, then in the previous example we saw that the equation of the tangent to graph(g) at $(0, 0)$ is $y = 0$; i.e., the x-axis is tangent to the curve. Note that in this case the tangent line crosses the curve at the point of tangency.

If $h(x) = x$ then for all $a \in \mathbf{R}$,

$$\lim_{x \to a} \frac{h(x) - h(a)}{x - a} = \lim_{x \to a} \frac{x - a}{x - a} = \lim_{x \to a} 1 = 1.$$

The equation of the tangent line to graph(h) at (a,a) is

$$y = a + 1(x-a)$$

or $y = x$.

Thus at each point on the curve the tangent line coincides with the curve.

Let $k(x) = (x^2)^{1/3}$. This is not the same as the function $l(x) = x^{2/3}$ since the domain of l is $\mathbf{R}_{\geq 0}$ while the domain of k is $\mathbf{R}$. (For all $x \in \mathbf{R}$ we have $x^2 \in \mathbf{R}_{\geq 0} = \text{dom}(g)$ where $g(x) = x^{1/3}$.)

I want to investigate $\displaystyle\lim_{x\to 0} \frac{k(x)-k(0)}{x-0} = \lim_{x\to 0} \frac{k(x)}{x}$. From the picture, I expect this graph to have an infinite slope at $(0,0)$, which means according to our definition that there is no tangent line at $(0,0)$. Let $\{x_n\} = \left\{\frac{1}{n^3}\right\}$. Then $\{x_n\} \to 0$, but

$$\frac{k(x_n)}{x_n} = \frac{\left(\frac{1}{n^6}\right)^{1/3}}{\left(\frac{1}{n^3}\right)} = \frac{\frac{1}{n^2}}{\frac{1}{n^3}} = n$$

so $\lim \left\{\frac{k(x_n)}{x_n}\right\}$ does not exist and hence $\displaystyle\lim_{x\to 0} \frac{k(x)}{x}$ does not exist.

10.23 Example. Let $f(x) = \sqrt{x}$ for $x \in \mathbf{R}_{\geq 0}$. Let $a \in \mathbf{R}^+$ and let $x \in \text{dom}(f) \setminus \{a\}$. Then

$$\begin{aligned} \frac{f(x)-f(a)}{x-a} &= \frac{\sqrt{x}-\sqrt{a}}{x-a} = \frac{\sqrt{x}-\sqrt{a}}{(\sqrt{x})^2-(\sqrt{a})^2} \\ &= \frac{\sqrt{x}-\sqrt{a}}{(\sqrt{x}-\sqrt{a})(\sqrt{x}+\sqrt{a})} = \frac{1}{\sqrt{x}+\sqrt{a}}. \end{aligned}$$

Hence

$$f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{x\to a} \frac{1}{\sqrt{x}+\sqrt{a}} = \frac{1}{\sqrt{a}+\sqrt{a}} = \frac{1}{2\sqrt{a}}; \qquad (10.24)$$

i.e.,

$$f'(a) = \frac{1}{2\sqrt{a}} \text{ for all } a \in \mathbf{R}^+.$$

In line (10.24) I used the fact that $\displaystyle\lim_{x\to a} \sqrt{x} = \sqrt{a}$, together with the sum and quotient rules for limits.

10.25 Exercise. Let $f(x) = \frac{1}{x}$. Sketch the graph of f. For what values of x do you expect $f'(x)$ to be -1? For what values of x do you expect $f'(x)$ to be positive? What do you expect to happen to $f'(x)$ when x is a small positive number? What do you expect to happen to $f'(x)$ when x is a small negative number?

Calculate $f'(a)$ for arbitrary $a \in \text{dom}(f)$. Does your answer agree with your prediction?

10.26 Exercise. Let $f(x) = \sin(x)$ for $-\pi < x < 4\pi$. Sketch the graph of f. Use the same scale on the x-axis and the y-axis.

On what intervals do you expect $f'(x)$ to be positive? On what intervals do you expect $f'(x)$ to be negative? Calculate $f'(0)$.

On the basis of symmetry, what do you expect to be the values of $f'(\pi)$, $f'(2\pi)$ and $f'(3\pi)$? For what x do you expect $f'(x)$ to be zero? On the basis of your guesses and your calculated value of $f'(0)$, draw a graph of f', where f' is the function that assigns $f'(x)$ to a generic number x in $(-\pi, 4\pi)$. On the basis of your graph, guess a formula for $f'(x)$.
(Optional) Prove that your guess is correct. (Some trigonometric identities will be needed.)

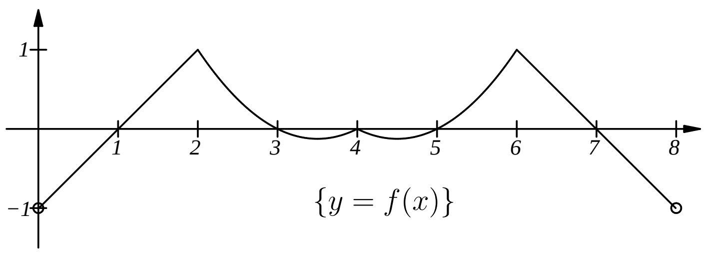
10.27 Exercise. Calculate $f'(x)$ if $f(x) = \frac{x}{x+1}$.

10.28 Exercise.

a) Find $f'(x)$ if $f(x) = x^2 - 2x$.

b) Find the equations for all the tangent lines to graph(f) that pass through the point $(0, -4)$. Make a sketch of graph(f) and the tangent lines.

10.29 Exercise. Consider the function $f\colon (0, 8) \to \mathbf{R}$ whose graph is shown below.



For what x in $(0,8)$ does $f'(x)$ exist? Sketch the graphs of f and f' on the same set of axes.

The following definition which involves time and motion and particles is not a part of our official development and will not be used for proving any theorems.

10.30 Definition (Velocity.) Let a particle $\mathbf{p}$ move on a number line in such a way that its coordinate at time t is $x(t)$, for all t in some interval J. (Here time is thought of as being specified by a number.) If t_0, t_1 are points in J with $t_0 < t_1$, then the *average velocity* of $\mathbf{p}$ for the time interval $[t_0, t_1]$ is defined to be

$$\frac{x(t_1) - x(t_0)}{t_1 - t_0} = \frac{\text{change in position}}{\text{change in time}}.$$

Note that $x(t_1) - x(t_0)$ is not necessarily the same as the distance moved in the time interval $[t_0, t_1]$. For example, if $x(t) = t(1-t)$ then $x(1) - x(0) = 0$, but the distance moved by $\mathbf{p}$ in the time interval $[0,1]$ is $\frac{1}{2}$. (The particle moves from 0 to $\frac{1}{4}$ at time $t = \frac{1}{2}$, and then back to 0.)

The *instantaneous velocity* of $\mathbf{p}$ at a time $t_0 \in J$ is defined to be

$$\lim_{t \to t_0} \frac{x(t) - x(t_0)}{t - t_0} = x'(t_0)$$

provided this limit exists. (If the limit does not exist, then the instantaneous velocity of $\mathbf{p}$ at t_0 is not defined.) If we draw the graph of the function x; i.e., $\{\big(t, x(t)\big) : t \in J\}$, then the velocity of $\mathbf{p}$ at time t_0 is by definition $x'(t_0) =$ slope of tangent to graph(x) at $\big(t_0, x(t_0)\big)$.

In applications we will usually express velocity in units like $\frac{\text{miles}}{\text{hour}}$. We will wait until we have developed some techniques for differentiation before we do any velocity problems.

The definition of velocity just given would have made no sense to Euclid or Aristotle. The Greek theory of proportion does not allow one to divide a length by a time, and Aristotle would no more divide a length by a time than he would add them. Question: Why is it that today in physics you are allowed to divide a length by a time, but you are not allowed to add a length to a time?

In Newton's calculus, the notion of instantaneous velocity or *fluxion* was taken as an undefined, intuitively understood concept, and the fluxions were calculated using methods similar to that used in the example on page 222.

The first "rigorous" definitions of limit of a function were given around 1820 by Bernard Bolzano (1781-1848) and Augustin Cauchy (1789-1857)[23, chapter 1]. The definition of limit of a function in terms of limits of sequences was given by Eduard Heine in 1872.

Chapter 11

Calculation of Derivatives

11.1 Derivatives of Some Special Functions

11.1 Theorem (Derivative of power functions.) *Let $r \in \mathbf{Q}$ and let $f(x) = x^r$. Here*

$$\text{domain}(f) = \begin{cases} \mathbf{R} & \text{if } r \in \mathbf{Z}_{\geq 0} \\ \mathbf{R} \setminus \{0\} & \text{if } r \in \mathbf{Z}^- \\ \mathbf{R}_{\geq 0} & \text{if } r \in \mathbf{Q}^+ \setminus \mathbf{Z} \\ \mathbf{R}^+ & \text{if } r \in \mathbf{Q}^- \setminus \mathbf{Z}. \end{cases}$$

Let a be an interior point of domain(f). Then f is differentiable at a, and

$$f'(a) = ra^{r-1}.$$

If $r = 0$ and $a = 0$ we interpret ra^{r-1} to be 0.

Proof: First consider the case $a \neq 0$. For all x in domain$(f) \setminus \{a\}$ we have

$$\frac{f(x) - f(a)}{x - a} = \frac{x^r - a^r}{x - a} = \frac{a^r\Big((\frac{x}{a})^r - 1\Big)}{a\Big((\frac{x}{a}) - 1\Big)} = a^{r-1}\frac{\Big((\frac{x}{a})^r - 1\Big)}{(\frac{x}{a} - 1)}.$$

Let $\{x_n\}$ be a generic sequence in domain$(f) \setminus \{a\}$ such that $\{x_n\} \to a$. Let $y_n = \dfrac{x_n}{a}$. Then $\{y_n\} \to 1$ and hence by theorem 7.10 we have $\left\{\dfrac{y_n^r - 1}{y_n - 1}\right\} \to r$ and hence

$$\left\{\frac{f(x_n) - f(a)}{x_n - a}\right\} = \left\{\frac{y_n^r - 1}{y_n - 1}\right\} \cdot a^{r-1} \to ra^{r-1}.$$

This proves the theorem in the case $a \neq 0$. If $a = 0$ then $r \in \mathbf{Z}_{\geq 0}$ (since for other values of r, 0 is not an interior point of domain(f)). In this case

$$\frac{f(x)-f(0)}{x-0} = \frac{x^r - 0^r}{x} = \begin{cases} 0 & \text{if } r=0 \text{ (remember } 0^0 = 1). \\ x^{r-1} & \text{if } r \neq 0. \end{cases}$$

Hence

$$f'(0) = \lim_{x \to 0} \frac{f(x)-f(0)}{x} = \begin{cases} 0 & \text{if } r=0, \\ 1 & \text{if } r=1, \\ 0 & \text{if } r>1. \end{cases}$$

Thus in all cases the formula $f'(x) = rx^{r-1}$ holds. |||

11.2 Corollary (Of the proof of theorem 11.1) *For all $r \in \mathbf{Q}$,*

$$\lim_{x \to 1} \frac{x^r - 1}{x-1} = r.$$

11.3 Theorem (Derivatives of sin **and** cos.**)** *Let $r \in \mathbf{R}$ and let $f(x) = \sin(rx)$, $g(x) = \cos(rx)$ for all $x \in \mathbf{R}$. Then f and g are differentiable on $\mathbf{R}$, and for all $x \in \mathbf{R}$*

$$f'(x) = r\cos(rx), \tag{11.4}$$
$$g'(x) = -r\sin(rx). \tag{11.5}$$

Proof: If $r = 0$ the result is clear, so we assume $r \neq 0$. For all $x \in \mathbf{R}$ and all $t \in \mathbf{R} \setminus \{x\}$, we have

$$\begin{aligned} \frac{\sin(rt)-\sin(rx)}{t-x} &= \frac{2\cos\left(\frac{r(t+x)}{2}\right)\sin\left(\frac{r(t-x)}{2}\right)}{t-x} \\ &= r\cos\left(\frac{r(t+x)}{2}\right) \cdot \frac{\sin\left(\frac{r(t-x)}{2}\right)}{\left(\frac{r(t-x)}{2}\right)}. \end{aligned}$$

(Here I've used an identity from theorem 9.21.) Let $\{x_n\}$ be a generic sequence in $\mathbf{R} \setminus \{x\}$ such that $\{x_n\} \to x$. Let $y_n = \dfrac{r(x_n+x)}{2}$ and let $z_n = \dfrac{r(x_n-x)}{2}$. Then $\{y_n\} \to rx$ so by lemma 9.34 we have $\{\cos(y_n)\} \to \cos(rx)$. Also $\{z_n\} \to 0$, and $z_n \in \mathbf{R} \setminus \{0\}$ for all $n \in \mathbf{Z}^+$, so by (9.38), $\left\{\dfrac{\sin(z_n)}{z_n}\right\} \to 1$. Hence

$$\left\{\frac{\sin(rx_n)-\sin(rx)}{x_n-x}\right\} = \left\{r\cos(y_n) \cdot \frac{\sin(z_n)}{z_n}\right\} \to r\cos(rx),$$

and this proves formula (11.4). ‖

The proof of (11.5) is similar.

11.6 Exercise. Prove that if $g(x) = \cos(rx)$, then $g'(x) = -r\sin(rx)$.

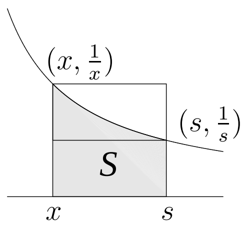
11.7 Theorem (Derivative of the logarithm.) *The logarithm function is differentiable on* $\mathbf{R}^+$*, and*

$$\ln'(x) = \frac{1}{x} \text{ for all } x \in \mathbf{R}^+.$$

Proof: Let $x \in \mathbf{R}^+$, and let $s \in \mathbf{R}^+ \setminus \{x\}$. Then

$$\frac{\ln(s) - \ln(x)}{s - x} = \frac{1}{s-x}\int_x^s \frac{1}{t}dt = \frac{1}{s-x}A_x^s\Big[\frac{1}{t}\Big].$$

Case 1: If $s > x$ then $A_x^s\Big[\frac{1}{t}\Big]$ represents the area of the shaded region S in the figure.



We have

$$B(x, s\colon 0, \frac{1}{s}) \subset S \subset B(x, s\colon 0, \frac{1}{x})$$

so by monotonicity of area

$$\frac{s-x}{s} \le A_x^s\Big[\frac{1}{t}\Big] \le \frac{s-x}{x}.$$

Thus

$$\frac{1}{s} \le \frac{1}{s-x}\int_x^s \frac{1}{t}dt \le \frac{1}{x}. \qquad (11.8)$$

Case 2. If $s < x$ we can reverse the roles of s and x in equation (11.8) to get

$$\frac{1}{x} \le \frac{1}{x-s}\int_s^x \frac{1}{t}dt \le \frac{1}{s}$$

or

$$\frac{1}{x} \leq \frac{1}{s-x}\int_x^s \frac{1}{t}dt \leq \frac{1}{s}.$$

In both cases it follows that

$$0 \leq \Big|\frac{1}{s-x}\int_x^s \frac{1}{t}dt - \frac{1}{x}\Big| \leq \Big|\frac{1}{s} - \frac{1}{x}\Big|.$$

Let $\{x_n\}$ be a generic sequence in $\mathbf{R}^+ \setminus \{x\}$ such that $\{x_n\} \to x$. Then $\Big\{\frac{1}{x_n} - \frac{1}{x}\Big\} \to 0$, so by the squeezing rule

$$\Big\{\frac{1}{x_n - x}\int_x^{x_n} \frac{1}{t}dt - \frac{1}{x}\Big\} \to 0,$$

i.e.

$$\left\{\frac{\ln(x_n) - \ln(x)}{x_n - x} - \frac{1}{x}\right\} \to 0.$$

Hence

$$\Big\{\frac{\ln(x_n) - \ln(x)}{x_n - x}\Big\} \to \frac{1}{x}.$$

We have proved that $\ln'(x) = \frac{1}{x}$. |||

11.9 Assumption (Localization rule for derivatives.) *Let f, g be two real valued functions. Suppose there is some $\epsilon \in \mathbf{R}^+$ and $a \in \mathbf{R}$ such that*

$$(a - \epsilon, a + \epsilon) \subset \text{domain}(f) \cap \text{domain}(g)$$

and such that

$$f(x) = g(x) \text{ for all } x \in (a - \epsilon, a + \epsilon).$$

If f is differentiable at a, then g is differentiable at a and $g'(a) = f'(a)$.

This is another assumption that is really a theorem, i.e. it can be proved. Intuitively this assumption is very plausible. It says that if two functions agree on an entire interval centered at a, then their graphs have the same tangents at a.

11.10 Theorem (Derivative of absolute value.) *Let $f(x) = |x|$ for all $x \in \mathbf{R}$. Then $f'(x) = \frac{x}{|x|}$ for all $x \in \mathbf{R} \setminus \{0\}$ and $f'(0)$ is not defined.*

Proof: Since

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ -x & \text{if } x < 0, \end{cases}$$

it follows from the localization theorem that

$$f'(x) = \begin{cases} 1 = \frac{x}{|x|} & \text{if } x > 0, \\ -1 = \frac{x}{|x|} & \text{if } x < 0. \end{cases}$$

To see that f is not differentable at 0, we want to show that

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

does not exist. Let $x_n = \dfrac{(-1)^n}{n}$. Then $\{x_n\} \to 0$, but $\dfrac{|x_n|}{x_n} = \dfrac{\left|\frac{(-1)^n}{n}\right|}{\frac{(-1)^n}{n}} = (-1)^n$ and we know that $\lim\{(-1)^n\}$ does not exist. Hence $\lim \dfrac{f(t) - f(0)}{t - 0}$ does not exist, i.e., f is not differentiable at 0.

11.11 Definition ($\frac{d}{dx}$ notation for derivatives.) An alternate notation for representing derivatives is:

$$\frac{d}{dx} f(x) = f'(x)$$

or

$$\frac{df}{dx} = f'(x).$$

This notation is used in the following way

$$\begin{aligned} \frac{d}{dx}\Big(\sin(6x)\Big) &= 6\cos(6x), \\ \frac{d}{dt}\Big(\cos(\frac{t}{3})\Big) &= -\frac{1}{3}\sin(\frac{t}{3}). \end{aligned}$$

Or:

Let $f = x^{1/2}$. Then $\dfrac{df}{dx} = \dfrac{1}{2}x^{-1/2}$.

Let $g(x) = \dfrac{1}{x}$. Then $\dfrac{dg}{dx} = \dfrac{d}{dx}\Big(g(x)\Big) = \dfrac{d}{dx}\Big(\dfrac{1}{x}\Big) = -\dfrac{1}{x^2}$.

The $\dfrac{d}{dx}$ notation is due to Leibnitz, and is older than our concept of function.

Leibnitz wrote the differentiation formulas as "$dx^a = ax^{a-1}dx$," or if $y = x^a$, then "$dy = ax^{a-1}dx$." The notation $f'(x)$ for derivatives is due to Joseph Louis Lagrange (1736-1813). Lagrange called $f'(x)$ the *derived function* of $f(x)$ and it is from this that we get our word *derivative*. Leibnitz called derivatives, *differentials* and Newton called them *fluxions*.

Many of the early users of the calculus thought of the derivative as the quotient of two numbers

$$\frac{df}{dx} = \frac{\text{difference in } f}{\text{difference in } x} = \frac{f(x) - f(t)}{x - t}$$

when $dx = x - t$ was "infinitely small". Today "infinitely small" real numbers are out of fashion, but some attempts are being made to bring them back. Cf *Surreal Numbers : How two ex-students turned on to pure mathematics and found total happiness : a mathematical novelette*, by D. E. Knuth.[30]. or *The Hyperreal Line* by H. Jerome Keisler[28, pp 207-237].

11.2 Some General Differentiation Theorems.

11.12 Theorem (Sum rule for derivatives.) *Let f, g be real valued functions with domain$(f) \subset \mathbf{R}$ and domain$(g) \subset \mathbf{R}$, and let $c \in \mathbf{R}$. Suppose f and g are differentiable at a. Then $f + g$, $f - g$ and cf are differentiable at a, and*

$$\begin{aligned} (f+g)'(a) &= f'(a) + g'(a) \\ (f-g)'(a) &= f'(a) - g'(a) \\ (cf)'(a) &= c \cdot f'(a). \end{aligned}$$

Proof: We will prove only the first statement. The proofs of the other statements are similar. For all $x \in \text{dom}(f)$ we have

$$\begin{aligned} \frac{(f+g)(x) - (f+g)(a)}{x - a} &= \frac{f(x) + g(x) - (f(a) + g(a))}{x - a} \\ &= \frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a} \end{aligned}$$

By the sum rule for limits of functions, it follows that

$$\lim_{x \to a}\left(\frac{(f+g)(x) - (f+g)(a)}{x - a}\right) = \lim_{x \to a}\left(\frac{f(x) - f(a)}{x - a}\right) + \lim_{x \to a}\left(\frac{g(x) - g(a)}{x - a}\right),$$

i.e.

$$(f+g)'(a) = f'(a) + g'(a). \;\|$$

11.13 Examples. If

$$f(x) = 27x^3 + \frac{1}{3x} + \sqrt{8x},$$

then

$$f(x) = 27x^3 + \frac{1}{3}x^{-1} + \sqrt{8} \cdot x^{1/2},$$

so

$$\begin{aligned} f'(x) &= 27 \cdot (3x^2) + \frac{1}{3}(-1 \cdot x^{-2}) + \sqrt{8} \cdot (\frac{1}{2}x^{-1/2}) \\ &= 81x^2 - \frac{1}{3x^2} + \sqrt{\frac{2}{x}}. \end{aligned}$$

If $g(x) = (3x^2+7)^2$, then $g(x) = 9x^4 + 42x^2 + 49$, so

$$g'(x) = 9 \cdot 4x^3 + 42 \cdot 2x = 36x^3 + 84x.$$

If $h(x) = \sin(4x) + \sin^2(4x)$, then $h(x) = \sin(4x) + \frac{1}{2}\Big(1 - \cos(8x)\Big)$, so

$$\begin{aligned} h'(x) &= 4\cos(4x) + \frac{1}{2}(-1)\Big(-8 \cdot \sin(8x)\Big) \\ &= 4\cos(4x) + 4\sin(8x). \end{aligned}$$

$$\frac{d}{ds}\Big(8\sin(4s) + s^2 + 4\Big) = 32\cos(4s) + 2s.$$

11.14 Exercise. Calculate the derivatives of the following functions:

a) $f(x) = (x^2+4x)^2$

b) $g(x) = \sqrt{3x^3} + \dfrac{4}{x^4}$

c) $h(t) = \ln(t) + \ln(t^2) + \ln(t^3)$

d) $k(x) = \ln(10 \cdot x^{5/2})$

e) $l(x) = 3\cos(x) + \cos(3x)$

f) $m(x) = \cos(x)\cos(3x)$

g) $n(x) = \Big(\sin^2(x) + \cos^2(x)\Big)^4$

11.15 Exercise. Calculate

a) $\dfrac{d}{dt}\Big(1 + t + \dfrac{t^2}{2!} + \dfrac{t^3}{3!} + \dfrac{t^4}{4!}\Big)$

b) $\dfrac{d}{dt}\Big(h_0 + v_0(t - t_0) - \dfrac{1}{2}g(t - t_0)^2\Big)$. Here h_0, v_0, t_0 and g are all constants.

c) $\dfrac{d}{dt}(|-100t|)$

11.16 Theorem (The product rule for derivatives.) *Let f and g be real valued functions with $\text{dom}(f) \subset \mathbf{R}$ and $\text{dom}(g) \subset \mathbf{R}$. Suppose f and g are both differentiable at a. Then fg is differentiable at a and*

$$(fg)'(a) = f(a) \cdot g'(a) + f'(a) \cdot g(a).$$

In particular, if $f = c$ is a constant function, we have

$$(cf)'(a) = c \cdot f'(a).$$

Proof: Let x be a generic point of $\text{dom}(f) \cap \text{dom}(g) \setminus \{a\}$. Then

$$\begin{aligned}\frac{(fg)(x) - (fg)(a)}{x - a} &= \frac{f(x)\Big(g(x) - g(a)\Big) + \Big(f(x) - f(a)\Big)g(a)}{x - a} \\ &= f(x)\left(\frac{g(x) - g(a)}{x - a}\right) + \left(\frac{f(x) - f(a)}{x - a}\right)g(a).\end{aligned}$$

We know that $\lim_{x \to a}\Big(\dfrac{g(x) - g(a)}{x - a}\Big) = g'(a)$ and $\lim_{x \to a}\Big(\dfrac{f(x) - f(a)}{x - a}\Big) = f'(a)$. If we also knew that $\lim_{x \to a} f(x) = f(a)$, then by basic properties of limits we could say that

$$(fg)'(a) = \lim_{x \to a}\frac{(fg)(x) - (fg)(a)}{x - a} = f(a)g'(a) + f'(a)g(a)$$

which is what we claimed.

This missing result will be needed in some other theorems, so I've isolated it in the following lemma.

11.17 Lemma (Differentiable functions are continuous.) *Let f be a real valued function such that* $\text{dom}(f) \subset \mathbf{R}^+$. *Suppose f is differentiable at a point $a \in$* $\text{dom}(f)$. *Then* $\lim\limits_{x \to a} f(x) = f(a)$. *(We will define "continuous" later. Note that neither the statement nor the proof of this lemma use the word "continuous" in spite of the name of the lemma.)*

Proof:

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

Hence by the product and sum rules for limits,

$$\lim_{x \to a} f(x) = f'(a) \cdot (a - a) + f(a) = f(a). \ \|$$

11.18 Example (Leibniz's proof of the product rule.) Leibniz stated the product rule as

$$dxy = xdy + ydx [34, page 143]^1$$

His proof is as follows:

> dxy is the difference between two successive xy's; let one of these be xy and the other $x + dx$ into $y + dy$; then we have
>
> $$dxy = \overline{\overline{x + dx} \cdot \overline{y + dy}} - xy = xdy + ydx + dxdy;$$
>
> the omission of the quantity $dxdy$ which is infinitely small in comparison with the rest, for it is supposed that dx and dy are infinitely small (because the lines are understood to be continuously increasing or decreasing by very small increments throughout the series of terms), will leave $xdy + ydx$.[34, page 143]

Notice that for Leibniz, the important thing is not the *derivative*, $\dfrac{dxy}{dt}$, but the infinitely small *differential*, dxy.

[1] The actual statement is $dxy = xdx + ydy$, but this is a typographical error, since the proof gives the correct formula.

11.19 Theorem (Derivative of a reciprocal.) *Let f be a real valued function such that* $\text{dom}(f) \subset \mathbf{R}$. *Suppose f is differentiable at some point a, and $f(a) \neq 0$. Then $\frac{1}{f}$ is differentiable at a, and*

$$\Big(\frac{1}{f}\Big)'(a) = \frac{-f'(a)}{\big(f(a)\big)^2}.$$

Proof: For all $x \in \text{dom}\Big(\frac{1}{f}\Big) \setminus \{a\}$

$$\frac{\frac{1}{f(x)} - \frac{1}{f(a)}}{x-a} = \frac{f(a) - f(x)}{(x-a)f(x)f(a)} = -\frac{\big(f(x) - f(a)\big)}{(x-a)} \cdot \frac{1}{f(x)f(a)}.$$

It follows from the standard limit rules that

$$\lim_{x \to a} \frac{\frac{1}{f(x)} - \frac{1}{f(a)}}{x-a} = -f'(a) \cdot \frac{1}{\big(f(a)\big)^2}.$$

11.20 Theorem (Quotient rule for derivatives.) *Let f, g be real valued functions with* $\text{dom}(f) \subset \mathbf{R}$ *and* $\text{dom}(g) \subset \mathbf{R}$. *Suppose f and g are both differentiable at a, and that $g(a) \neq 0$. Then $\frac{f}{g}$ is differentiable at a, and*

$$\Big(\frac{f}{g}\Big)'(a) = \frac{g(a)f'(a) - f(a)g'(a)}{\big(g(a)\big)^2}.$$

11.21 Exercise. Prove the quotient rule.

11.22 Examples. Let

$$f(x) = \frac{\sin(x)}{x} \text{ for } x \in \mathbf{R} \setminus \{0\}.$$

Then by the quotient rule

$$f'(x) = \frac{x(\cos(x)) - \sin(x)}{x^2}.$$

Let $h(x) = x^2 \cdot |x|$. Then by the product rule

$$h'(x) = x^2\Big(\frac{x}{|x|}\Big) + 2x|x| = x|x| + 2x|x| = 3x|x|$$

(since $\dfrac{x^2}{|x|} = \dfrac{|x|^2}{|x|} = |x|$).

The calculation is not valid at $x = 0$ (since $|x|$ is not differentiable at 0, and we divided by $|x|$ in the calculation. However h is differentiable at 0 since $\lim\limits_{t\to 0}\dfrac{h(t)-h(0)}{t-0} = \lim\limits_{t\to 0}\dfrac{t^2|t|}{t} = \lim\limits_{t\to 0} t|t| = 0$, i.e., $h'(0) = 0 = 3\cdot 0\cdot|0|$. Hence the formula

$$\frac{d}{dx}(x^2|x|) = 3x|x|$$

is valid for all $x \in \mathbf{R}$.

Let $g(x) = \ln(x)\cdot\sin(10x)\cdot\sqrt{x}$. Consider g to be a product $g = hk$ where $h(x) = \ln(x)\cdot\sin(10x)$ and $k(x) = \sqrt{x}$. Then we can apply the product rule twice to get

$$\begin{aligned} g'(x) &= \Big(\ln(x)\cdot\sin(10x)\Big)\cdot\frac{1}{2\sqrt{x}} \\ &\quad+\Bigg(\ln(x)\cdot\Big(10\cos(10x)\Big) + \frac{1}{x}\sin(10x)\Bigg)\sqrt{x}. \end{aligned}$$

11.23 Exercise (Derivatives of tangent, cotangent, secant, cosecant.) We define functions tan, cot, sec, and csc by

$$\tan(x) = \frac{\sin(x)}{\cos(x)},\quad \cot(x) = \frac{\cos(x)}{\sin(x)},$$

$$\sec(x) = \frac{1}{\cos(x)},\quad \csc(x) = \frac{1}{\sin(x)}.$$

The domains of these functions are determined by the definition of the domain of a quotient, e.g. $\text{dom}(\sec) = \{x \in \mathbf{R}\colon \cos x \neq 0\}$. Prove that

$$\begin{aligned} &\tfrac{d}{dx}\tan(x) = \sec^2(x), && \tfrac{d}{dx}\cot(x) = -\csc^2 x, \\ &\tfrac{d}{dx}\sec(x) = \tan(x)\sec(x), && \tfrac{d}{dx}\csc(x) = -\cot(x)\csc(x). \end{aligned}$$

(You should memorize these formulas. Although they are easy to derive, later we will want to use them backwards; i.e., we will want to find a function whose derivative is $\sec^2(x)$. It is not easy to derive the formulas backwards.)

11.24 Exercise. Calculate the derivatives of the following functions. Simplify your answers if you can.

a) $f(x) = x \cdot \ln(x) - x$.

b) $g(x) = \dfrac{ax+b}{cx+d}$ (here a, b, c, d are constants).

c) $k(x) = (x^2 + 3x + 10)(x^2 + 3x + 12)$.

d) $m(x) = \dfrac{\cos(6x)}{\cos(7x)}$.

11.25 Exercise. Let f, g, h, and k be differentiable functions defined on $\mathbf{R}$.
a) Express $(fgh)'$ in terms of f, f', g, g', h and h'.
b) On the basis of your answer for part a), try to guess a formula for $(fghk)'$. Then calculate $(fghk)'$, and see whether your guess was right.

11.3 Composition of Functions

11.26 Definition ($f \circ g$.) Let A, B, C, D be sets and let $f\colon A \to B,\ g\colon C \to D$ be functions. The composition of f and g is the function $f \circ g$ defined by:

$$\begin{aligned} \text{codomain}(f \circ g) &= B = \text{codomain}(f). \\ \text{dom}(f \circ g) &= \{x \in C\colon g(x) \in A\} \\ &= \{x \in \text{dom}(g)\colon g(x) \in \text{dom}(f)\}; \end{aligned}$$

i.e., $\text{dom}(f \circ g)$ is the set of all points x such that $f\big(g(x)\big)$ is defined. The rule for $f \circ g$ is

$$(f \circ g)(x) = f\big(g(x)\big) \text{ for all } x \in \text{dom}(f \circ g).$$

11.27 Example. If $f(x) = \sin(x)$ and $g(x) = x^2 - 2$, then

$$(f \circ g)(x) = \sin(x^2 - 2)$$

and

$$(g \circ f)(x) = \sin^2(x) - 2.$$

Thus

$$(f \circ g)(0) = \sin(-2) \text{ and } (g \circ f)(0) = -2 \neq (f \circ g)(0).$$

So in this case $f \circ g \neq g \circ f$. Thus composition is not a commutative operation.

If $h(x) = \ln(x)$ and $k(x) = |x|$, then

$$(h \circ k)(x) = \ln(|x|)$$

and

$$(k \circ h)(x) = |\ln(x)|.$$

11.28 Exercise. For each of the functions F below, find functions f and g such that $F = f \circ g$. Then find a formula for $g \circ f$.

a) $F(x) = \ln(\tan(x))$.

b) $F(x) = \sin(4(x^2 + 3))$.

c) $F(x) = |\sin(x)|$.

11.29 Exercise. Let

$$\begin{aligned} f(x) &= \sqrt{1 - x^2}, \\ g(x) &= \frac{1}{1 - x}. \end{aligned}$$

Calculate formulas for $f \circ f$, $f \circ (f \circ f)$, $(f \circ f) \circ f$, $g \circ g$, $(g \circ g) \circ g$, and $g \circ (g \circ g)$.

11.30 Entertainment (Composition problem.) From the previous exercise you should be able to find a subset A of $\mathbf{R}$, and a function $f : A \to \mathbf{R}$ such that $(f \circ f)(x) = x$ for all $x \in A$. You should also be able to find a subset B of $\mathbf{R}$ and a function $g : B \to \mathbf{R}$ such that $(g \circ (g \circ g))(x) = x$ for all $x \in B$. Can you find a subset C of $\mathbf{R}$, and a function $h : C \to \mathbf{R}$ such that $(h \circ (h \circ (h \circ h)))(x) = x$ for all $x \in C$? One obvious example is the function f from the previous example. To make the problem more interesting, also add the condition that $(h \circ h)(x) \neq x$ for some x in C.

11.31 Theorem (Chain rule.) *Let f, g be real valued functions such that* $\mathrm{dom}(f) \subset \mathbf{R}$ *and* $\mathrm{dom}(g) \subset \mathbf{R}$. *Suppose $a \in \mathrm{dom}(g)$ and $g(a) \in \mathrm{dom} f$, and g is differentiable at a and f is differentiable at $g(a)$. Then $f \circ g$ is differentiable at a, and*

$$(f \circ g)'(a) = f'\Big(g(a)\Big) \cdot g'(a).$$

Before we prove the theorem we will give a few examples of how it is used:

11.32 Example. Let $H(x) = \sqrt{10 + \sin x}$. Then $H = f \circ g$ where

$$f(x) = \sqrt{x}, \qquad g(x) = 10 + \sin(x),$$

$$f'(x) = \frac{1}{2\sqrt{x}}, \quad g'(x) = \cos(x).$$

Hence

$$\begin{aligned} H'(x) &= f'\Big(g(x)\Big) \cdot g'(x) \\ &= \frac{1}{2\sqrt{10 + \sin(x)}} \cdot \cos(x). \end{aligned}$$

Let $K(x) = \ln(5x^2 + 1)$. Then $K = f \circ g$ where

$$f(x) = \ln(x), \quad g(x) = 5x^2 + 1,$$

$$f'(x) = \frac{1}{x}, \qquad g'(x) = 10x.$$

Hence

$$\begin{aligned} K'(x) &= f'\Big(g(x)\Big) \cdot g'(x) \\ &= \frac{1}{5x^2 + 1} \cdot 10x = \frac{10x}{5x^2 + 1}. \end{aligned}$$

Usually I will not write out all of the details of a calculation like this. I will just write:

Let $f(x) = \tan(2x + 4)$. Then $f'(x) = \sec^2(2x + 4) \cdot 2$.

Proof of chain rule: Suppose g is differentiable at a and f is differentiable at $g(a)$. Then

$$\frac{f\big(g(x)\big) - f\big(g(a)\big)}{x - a} = \frac{f\big(g(x)\big) - f\big(g(a)\big)}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a}. \tag{11.33}$$

Since g is differentiable at a, we know that

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

Hence the theorem will follow from (11.33), the definition of derivative, and the product rule for limits of functions, if we can show that

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

Since g is differentiable at a, it follows from lemma 11.17 that

$$\lim_{x \to a} g(x) = g(a). \tag{11.34}$$

Let $\{x_n\}$ be a generic sequence in $\text{dom}(f \circ g) \setminus \{a\}$, such that $\{x_n\} \to a$. Then by (11.34), we have

$$\lim\{g(x_n)\} = g(a). \tag{11.35}$$

Since f is differentiable at $g(a)$, we have

$$\lim_{t \to g(a)} \frac{f(t) - f(g(a))}{t - g(a)} = f'(g(a)).$$

From this and (11.35) it follows that

$$\lim \left\{ \frac{f(g(x_n)) - f(g(a))}{g(x_n) - g(a)} \right\} = f'(g(a)).$$

Since this holds for a generic sequence $\{x_n\}$ in $\text{dom}(f \circ g) \setminus \{a\}$, we have

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

which is what we wanted to prove. To complete the proof, I should show that a is an interior point of $\text{dom}(f \circ g)$. This turns out to be rather tricky, so I will omit the proof.

Remark: Our proof of the chain rule is not valid in all cases, but it is valid in all cases where you are likely to use it. The proof fails in the case where every interval $\Big(g(a)-\epsilon,\ g(a)+\epsilon\Big)$ contains a point $b \neq a$ for which $g(b) = g(a)$. (You should check the proof to see where this assumption was made.) Constant functions g satisfy this condition, but if g is constant then $f \circ g$ is also constant so the chain rule holds trivially in this case. Since the proof in the general case is more technical than illuminating, I am going to omit it. Can you find a non-constant function g for which the proof fails?

11.36 Example. If f is differentiable at x, and $f(x) \neq 0$, then

$$\frac{d}{dx}\Big(|f(x)|\Big) = \frac{f(x)}{|f(x)|} f'(x).$$

Also

$$\begin{aligned} \frac{d}{dx}\left(\ln\Big(|f(x)|\Big)\right) &= \frac{1}{|f(x)|}\frac{d}{dx}\Big(|f(x)|\Big) \\ &= \frac{1}{|f(x)|}\frac{f(x)}{|f(x)|}f'(x) = \frac{f(x)f'(x)}{f(x)^2} = \frac{f'(x)}{f(x)}; \end{aligned}$$

i.e.,

$$\frac{d}{dx}\Big(\ln|f(x)|\Big) = \frac{f'(x)}{f(x)} \tag{11.37}$$

I will use this relation frequently.

11.38 Example (Logarithmic differentiation.) Let

$$h(x) = \frac{\sqrt{(x^2+1)}(x^2-4)^{10}}{(x^3+x+1)^3}. \tag{11.39}$$

The derivative of h can be found by using the quotient rule and the product rule and the chain rule. I will use a trick here which is frequently useful. I have

$$\ln\Big(|h(x)|\Big) = \frac{1}{2}\ln(x^2+1) + 10\ln(|x^2-4|) - 3\ln(|x^3+x+1|).$$

Now differentiate both sides of this equation using (11.37) to get

$$\frac{h'(x)}{h(x)} = \frac{1}{2}\frac{2x}{x^2+1} + 10\frac{2x}{x^2-4} - 3\frac{3x^2+1}{x^3+x+1}.$$

Multiply both sides of the equation by $h(x)$ to get

$$h'(x) = \frac{\sqrt{x^2+1}(x^2-4)^{10}}{(x^3+x+1)^3}\Big[\frac{x}{x^2+1} + \frac{20x}{x^2-4} - \frac{3(3x^2+1)}{x^3+x+1}\Big].$$

This formula is not valid at points where $h(x) = 0$, because we took logarithms in the calculation. Thus h is differentiable at $x = 2$, but our formula for $h'(x)$ is not defined when $x = 2$.

The process of calculating f' by first taking the logarithm of the absolute value of f and then differentiating the result, is called *logarithmic differentiation.*

11.40 Exercise. Let h be the function defined in (11.39) Show that h is differentiable at 2, and calculate $h'(2)$.

11.41 Exercise. Find derivatives for the functions below. (Assume here that f is a function that is differentiable at all points being considered.)

a) $F(x) = \sin(f(x))$.

b) $G(x) = \cos(f(x))$.

c) $H(x) = (f(x))^r$, where r is a rational number.

d) $K(x) = \ln((f(x))$.

e) $L(x) = |f(x)|$.

f) $M(x) = \tan(f(x))$.

g) $N(x) = \cot(f(x))$.

h) $P(x) = \sec(f(x))$.

i) $Q(x) = \csc(f(x))$.

j) $R(x) = \ln(|f(x)|)$.

11.42 Exercise. Find derivatives for the functions below. (Assume here that f is a function that is differentiable at all points being considered.)

a) $F(x) = f(\sin(x))$.

b) $G(x) = f(\cos(x))$.

c) $H(x) = f(x^r)$, where r is a rational number.

d) $K(x) = f(\ln(x))$.

e) $L(x) = f(|x|)$.

f) $M(x) = f(\tan(x))$.

g) $N(x) = f(\cot(x))$.

h) $P(x) = f(\sec(x))$.

i) $Q(x) = f(\csc(x))$.

j) $R(x) = f(\ln(|x|)$.

11.43 Exercise. Calculate the derivatives of the following functions. Simplify your answers.

a) $a(x) = \sin^3(x) = (\sin(x))^3$.

b) $b(x) = \sin(x^3)$.

c) $c(x) = (x^2 + 4)^{10}$.

d) $f(x) = \sin(4x^2 + 3x)$.

e) $g(x) = \ln\Big(|\cos(x)|\Big)$.

f) $h(x) = \ln\Big(|\sec(x)|\Big)$.

g) $k(x) = \ln\Big(|\sec(x) + \tan(x)|\Big)$.

h) $l(x) = \ln\Big(|\csc(x) + \cot(x)|\Big)$.

i) $m(x) = 3x^3 \ln(5x) - x^3$.

j) $n(x) = \sqrt{x^2+1} + \ln\Big(\frac{\sqrt{x^2+1} - 1}{x}\Big)$.

k) $p(x) = \frac{1}{2}(x+4)^2 - 8x + 16\ln(x+4)$.

l) $q(x) = \dfrac{x}{2}\Big[\sin\Big(\ln(|6x|)\Big) - \cos\Big(\ln(|6x|)\Big)\Big].$

Chapter 12

Extreme Values of Functions

12.1 Continuity

12.1 Definition (Continuity at a point.) Let f be a real valued function such that $\text{dom}(f) \subset \mathbf{R}$. Let $a \in \text{dom}(f)$. We say that f *is continuous at* a if and only if

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

Remark: According to this definition, in order for f to be continuous at a we must have

$$a \in \text{dom}(f)$$

and

$$a \text{ is approachable from } \text{dom}(f).$$

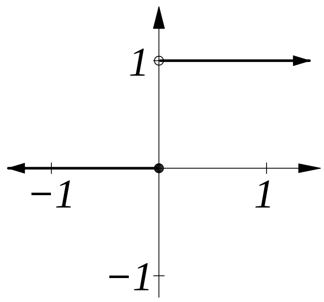
The second condition is often not included in the definition of continuity, so this definition does not quite correspond to the usual definition.
Remark: The method we will usually use to show that a function f is *not* continuous at a point a, is to find a sequence $\{x_n\}$ in $\text{dom}(f) \setminus \{a\}$ such that $\{x_n\} \to a$, but $\{f(x_n\}$ either diverges or converges to a value different from $f(a)$.

12.2 Definition (Continuity on a set.) Let f be a real valued function such that $\text{domain}(f) \subset \mathbf{R}$, and let S be a subset of $\text{domain}(f)$. We say that f *is continuous on* S if f is continuous at every point in S. We say that f *is continuous* if f is continuous at every point of $\text{domain}(f)$.

12.3 Example (sin, cos, ln and power functions are continuous.) We proved in lemma 11.17 that a function is continuous at every point at which it is differentiable. (You should now check the proof of that lemma to see that we did prove this.) Hence sin, cos, ln, and x^n (for $n \in \mathbf{Z}$) are all continuous on their domains, and if $r \in \mathbf{Q} \setminus \mathbf{Z}$, then x^r is continuous on $\mathbf{R}^+$.

12.4 Example. Let

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$



Then f is not continuous at 0. For the sequence $\{\frac{1}{n}\}$ converges to 0, but $\{f(\frac{1}{n})\} = \{1\} \to 1 \neq f(0)$.

Our limit rules all give rise to theorems about continuous functions.

12.5 Theorem (Properties of continuous functions.) *Let f, g be real valued functions with* $\text{dom}(f) \subset \mathbf{R}$, $\text{dom}(g) \subset \mathbf{R}$, *and let $c, a \in \mathbf{R}$. If f and g are continuous at a and if a is approachable from* $\text{dom}(f) \cap \text{dom}(g)$, *then $f + g$, $f - g$, fg, and cf are continuous at a. If in addition, $g(a) \neq 0$ then $\frac{f}{g}$ is also continuous at a.*

Proof: Suppose f and g are continuous at a, and a is approachable from $\text{dom}(f) \cap \text{dom}(g)$. Then

$$\lim_{x \to a} f(x) = f(a) \text{ and } \lim_{x \to a} g(x) = g(a).$$

By the sum rule for limits (theorem 10.15) it follows that

$$\begin{aligned} \lim_{x \to a}(f + g)(x) &= \lim_{x \to a} f(x) + \lim_{x \to a} g(x) \\ &= f(a) + g(a) = (f + g)(a). \end{aligned}$$

Thus $f+g$ is continuous at a. The proofs of the other parts of the theorem are similar.

12.6 Example (An everywhere discontinuous function.) Let D be the example of a non-integrable function defined in equation (8.37). Then D is not continuous at any point of $[0,1]$. Recall

$$D(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \neq S \end{cases}$$

where S is a subset of $[0,1]$ such that every subinterval of $[0,1]$ of positive length contains a point in S and a point not in S. Let $x \in [0,1]$.

Case 1. If $x \in S$ we can find a sequence of points $\{t_n\}$ in $[0,1] \setminus S$ such that $\{t_n\} \to x$. Then

$$\{D(t_n)\} = \{0\} \to 0 \neq D(x)$$

so D is not continuous at x.

Case 2. If $x \notin S$ we can find a sequence of points $\{s_n\}$ in S such that $\{s_n\} \to x$. Then

$$\{D(s_n)\} = \{1\} \to 1 \neq D(x)$$

so D is not continuous at x.

12.7 Example. Let

$$h(x) = \sqrt{x} \text{ for } x \in \mathbf{R}_{\geq 0}.$$

I claim that h is continuous. We know that h is differentiable on $\mathbf{R}^+$, so H is continuous at each point of $\mathbf{R}^+$. In example 10.13 we showed that $\lim\limits_{x \to 0} h(x) = 0 = h(0)$ so h is also continuous at 0.

12.8 Example. Let

$$\begin{aligned} f(x) &= -x^2, \\ g(x) &= \sqrt{x}. \end{aligned}$$

Then f and g are both continuous functions. Now

$$(g \circ f)(x) = \sqrt{-x^2}$$

and hence $g \circ f$ is not continuous. The domain of $g \circ f$ contains just one point, and that point is not approachable from $\text{dom}(g \circ f)$.

12.9 Theorem (Continuity of compositions.) *Let f, g be functions with domains contained in $\mathbf{R}$ and let $a \in \mathbf{R}$. Suppose that f is continuous at a and g is continuous at $f(a)$. Then $g \circ f$ is continuous at a,* provided that a is approachable from $\text{dom}(g \circ f)$.

Proof: Suppose f is continuous at a and g is continuous at $f(a)$, and a is approachable from $\text{dom}(g \circ f)$. Let $\{x_n\}$ be a sequence in $\text{dom}(g \circ f) \setminus \{a\}$ such that $\{x_n\} \to a$. Then $\{f(x_n)\} \to f(a)$ since f is continuous at a. Hence $\{g\big(f(x_n)\big)\} \to g\big(f(a)\big)$ since g is continuous at $f(a)$; i.e.,

$$\{(g \circ f)(x_n)\} \to (g \circ f)(a).$$

Hence $g \circ f$ is continuous at a.

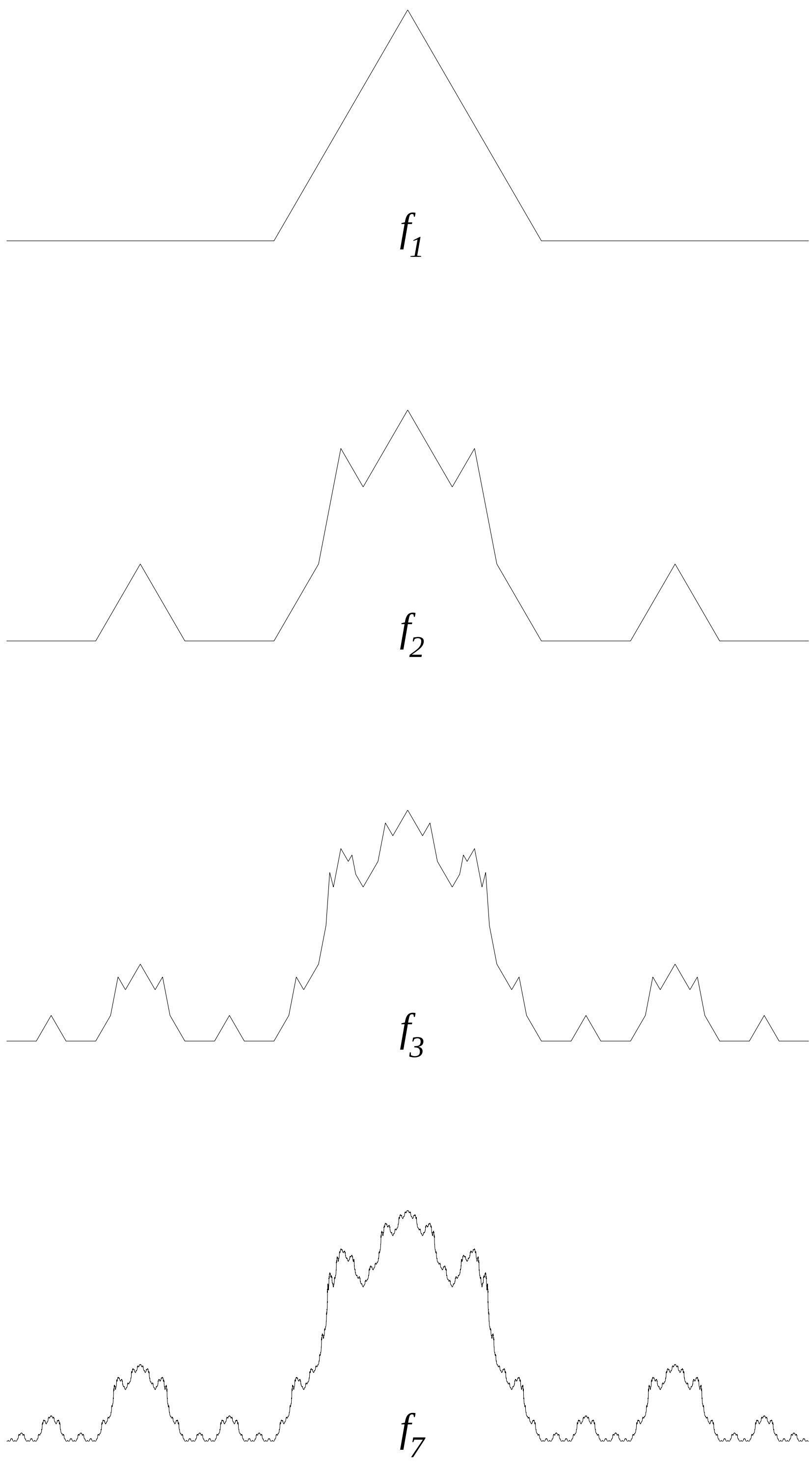
12.2 *A Nowhere Differentiable Continuous Function.

We will now give an example of a function f that is continuous at every point of $[0, 1]$ and differentiable at no point of $[0, 1]$. The first published example of such a function appeared in 1874 and was due to Karl Weierstrass(1815-1897) [29, page 976]. The example described below is due to Helga von Koch (1870-1924), and is a slightly modified version of Koch's snowflake. From the discussion in section 2.6, it is not really clear what we would mean by the *perimeter* of a snowflake, but it is pretty clear that whatever the perimeter might be, it is not the graph of a function. However, a slight modification of Koch's construction yields an everywhere continuous but nowhere differentiable *function.*

We will construct a sequence $\{f_n\}$ of functions on $[0, 1]$. The graph of f_n will be a polygonal line with 4^{n-1} segments. We set

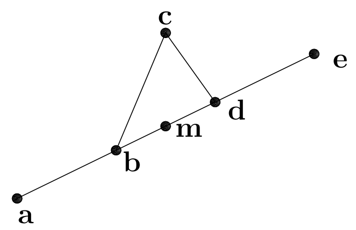
$$f_1(x) = 0 \quad \text{for} \quad 0 \le x \le 1$$

so that the graph of f_1 is the line segment from $(0, 0)$ to $(0, 1)$.



Approximations to a nowhere differentiable function

In general the graph of f_{n+1} is obtained from the graph of f_n by replacing each segment $[\mathbf{ae}]$ in the graph of f_n by four segments $[\mathbf{ab}], [\mathbf{bc}], [\mathbf{cd}]$, and $[\mathbf{de}]$ constructed according to the following three rules:



i) The points **b** and **d** trisect the segment $[\mathbf{ae}]$.

ii) The point **c** lies above the midpoint **m** of $[\mathbf{ae}]$.

iii) $\text{distance}(\mathbf{m}, \mathbf{c}) = \dfrac{\sqrt{3}}{2} \text{ distance}(\mathbf{b}, \mathbf{d})$.

The graphs of f_2, f_3, f_4 and f_7 are shown on page 260. It can be shown that for each $x \in [0, 1]$ the sequence $\{f_n(x)\}$ converges. Define f on $[0, 1]$ by

$$f(x) = \lim\{f_n(x)\} \text{ for all } x \in [0, 1].$$

It turns out that f is continuous on $[0, 1]$ and differentiable nowhere on $[0, 1]$. A proof of this can be found in [31, page 168].

The function f provides us with an example of a continuous function that is not piecewise monotonic over any interval.

12.3 Maxima and Minima

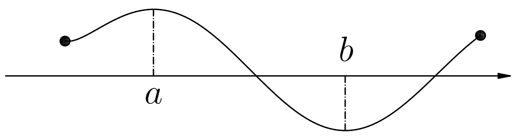
12.10 Definition (Maximum, minimum, extreme points.) Let A be a set, let $f: A \to \mathbf{R}$ and let $a \in A$. We say that f has a *maximum* at a if

$$f(a) \geq f(x) \text{ for all } x \in A,$$

and we say that f has a *minimum* at a if

$$f(a) \leq f(x) \text{ for all } x \in A.$$

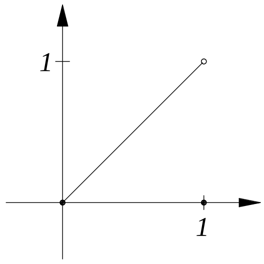
Points a where f has a maximum or a minimum are called *extreme points of* f.



f has a maximum at a and a minimum at b

12.11 Example. Let $f\colon [0,1] \to \mathbf{R}$ be defined by

$$f(x) = \begin{cases} x & \text{if } 0 \leq x < 1 \\ 0 & \text{if } x = 1. \end{cases} \tag{12.12}$$

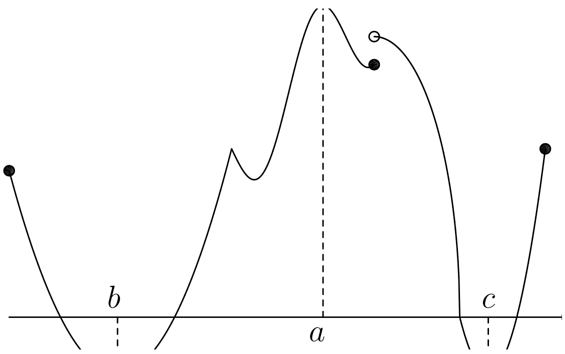


$\{y = f(x)\}$

Then f has a minimum at 0 and at 1, but f has no maximum. To see that f has no maximum, observe that if $a \in [0,1)$ then $\dfrac{1+a}{2} \in [0,1)$ and

$$f(\frac{1+a}{2}) = \frac{1+a}{2} > \frac{a+a}{2} = a = f(a).$$

If g is the function whose graph is shown, then g has a maximum at a, and g has minimums at b and c.



12.13 Assumption (Extreme value property.) If f is a continuous function on the interval $[a, b]$, then f has a maximum and a minimum on $[a, b]$.

The extreme value property is another assumption that is really a theorem, (although the proof requires yet another assumption, namely *completeness* of the real numbers.)

The following exercise shows that all of the hypotheses of the extreme value property are necessary.

12.14 Exercise.

a) Give an example of a continuous function f on $(0, 1)$ such that f has no maximum on $(0, 1)$.

b) Give an example of a bounded continuous function g on the closed interval $[0, \infty)$, such that g has no maximum on $[0, \infty)$

c) Give an example of a function h on $[0, 1]$ such that h has no maximum on $[0, 1]$.

d) Give an example of a continuous function k on $[0, \infty)$ that has neither a maximum nor a minimum on $[0, \infty)$, or else explain why no such function exists.

12.15 Exercise.

a) Show that every continuous function from an interval $[a, b]$ to $\mathbf{R}$ is bounded. (Hint: Use the extreme value property,)

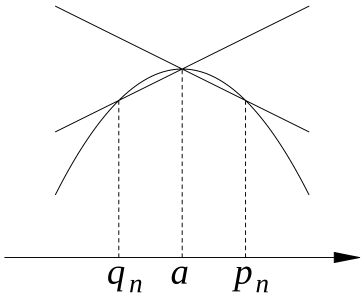
b) Is it true that every continuous function from an open interval (a, b) to $\mathbf{R}$ is bounded?

c) Give an example of a function from $[0, 1]$ to $\mathbf{R}$ that is not bounded.

12.16 Definition (Critical point, critical set.) Let f be a real valued function such that $\text{dom}(f) \subset \mathbf{R}$. A point $a \in \text{dom}(f)$ is called a *critical point for* f if $f'(a) = 0$. The set of critical points for f is the *critical set* for f. The points x in the critical set for f correspond to points $\big(x, f(x)\big)$ where the graph of f has a horizontal tangent.

12.17 Theorem (Critical point theorem I.) *Let f be a real valued function with* $\text{dom}(f) \subset \mathbf{R}$. *Let $a \in \mathbf{R}$. If f has a maximum (or a minimum) at a, and f is differentiable at a, then $f'(a) = 0$.*

Proof: We will consider only the case where f has a maximum. Suppose f has a maximum at a and f is differentiable at a. Then a is an interior point of $\text{dom}(f)$ so we can find sequences $\{p_n\}$ and $\{q_n\}$ in $\text{dom}(f) \setminus \{a\}$ such that $\{p_n\} \to a$, $\{q_n\} \to a$, $p_n > a$ for all $n \in \mathbf{Z}^+$, and $q_n < a$ for all $n \in \mathbf{Z}^+$.



Since f has a maximum at a, we have $f(p_n) - f(a) \leq 0$ and $f(q_n) - f(a) \leq 0$ for all n. Hence

$$\frac{f(p_n) - f(a)}{p_n - a} \leq 0 \text{ and } \frac{f(q_n) - f(a)}{q_n - a} \geq 0 \text{ for all } n.$$

Hence by the inequality theorem for limits,

$$f'(a) = \lim \left\{ \frac{f(p_n) - f(a)}{p_n - a} \right\} \leq 0 \text{ and } f'(a) = \lim \left\{ \frac{f(q_n) - f(a)}{q_n - a} \right\} \geq 0.$$

It follows that $f'(a) = 0$. |||

12.18 Definition (Local maximum and minimum.) Let f be a real valued function whose domain is a subset of $\mathbf{R}$. Let $a \in \text{dom}(f)$. We say that f has a *local maximum* at a if there is a positive number δ such that

$$f(a) \geq f(x) \text{ for all } x \in \text{dom}(f) \cap (a - \delta, a + \delta),$$

and we say that f has a *local minimum* at a if there is a positive number δ such that

$$f(a) \leq f(x) \text{ for all } x \in \text{dom}(f) \cap (a - \delta, a + \delta).$$

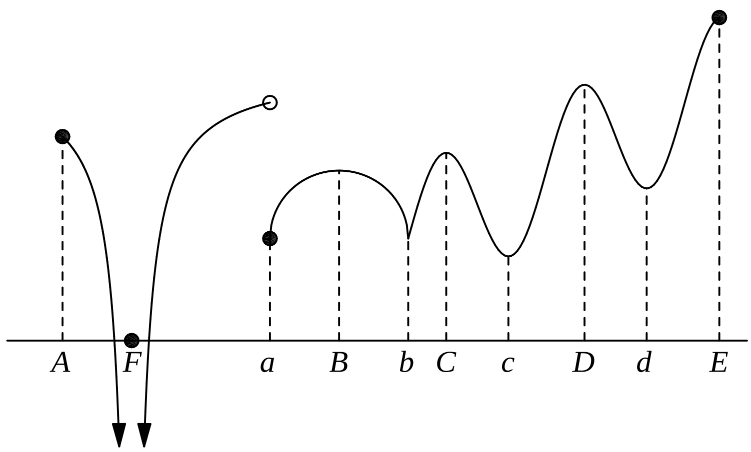
Sometimes we say that f has a *global maximum at* a to mean that f has a maximum at a, when we want to emphasize that we do not mean local maximum. If f has a local maximum or a local minimum at a we say f has a *local extreme point* at a.

12.19 Theorem (Critical point theorem II.) *Let f be a real valued function with* $\text{dom}(f) \subset \mathbf{R}$. *Let $a \in \mathbf{R}$. If f has a local maximum or minimum at a, and f is differentiable at a, then $f'(a) = 0$.*

Proof: The proof is the same as the proof of theorem 12.17.

12.20 Examples. If f has a maximum at a, then f has a local maximum at a.

The function g whose graph is shown in the figure has local maxima at A, B, C, D, E, F and local minima at a, b, c, and d. It has a global maximum at E, and it has no global minimum.



From the critical point theorem, it follows that to investigate the extreme points of f, we should look at critical points, or at points where f is not differentiable (including endpoints of domain f).

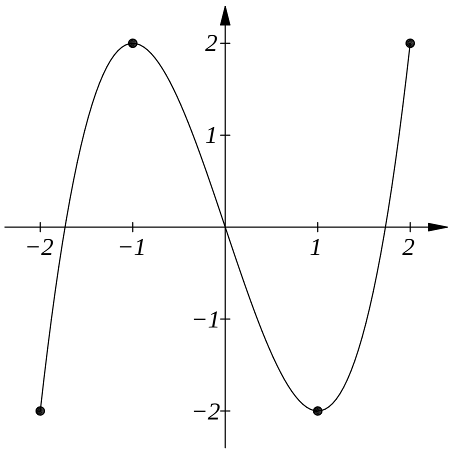
12.21 Example. Let $f(x) = x^3 - 3x$ for $-2 \leq x \leq 2$. Then f is differentiable everywhere on $\text{dom}(f)$ except at 2 and -2. Hence, any local extreme points are critical points of f or are in $\{2, -2\}$. Now

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

From this we see that the critical set for f is $\{-1, 1\}$. Since f is a continuous function on a closed interval $[-2, 2]$ we know that f has a maximum and a minimum on $[-2, 2]$. Now

$$f(-2) = -2,\ f(-1) = 2,\ f(1) = -2,\ f(2) = 2.$$

Hence f has global maxima at -1 and 2, and f has global minima at -2 and 1. The graph of f is shown.



$\{y = x^3 - 3x)\}$

12.22 Example. Let

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

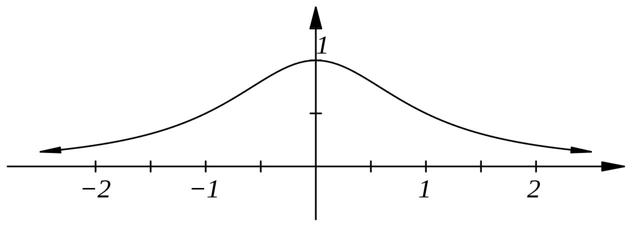
Here $\mathrm{dom}(f) = \mathbf{R}$ and clearly $f(x) > 0$ for all x. I can see by inspection that f has a maximum at 0; i.e.,

$$f(x) = \frac{1}{1 + x^2} \leq \frac{1}{1 + 0} = 1 = f(0) \text{ for all } x \in \mathbf{R}$$

I also see that $f(-x) = f(x)$, and that f is strictly decreasing on $\mathbf{R}^+$

$$0 < x < t \implies x^2 < t^2 \implies 1 + x^2 < 1 + t^2 \implies \frac{1}{1 + x^2} > \frac{1}{1 + t^2}$$

thus f has no local extreme points other than 0. Also $f(x)$ is very small when x is large. There is no point in calculating the critical points here because all the information about the extreme points is apparent without the calculation.



$\{y = \frac{1}{1+x^2}\}$

12.23 Exercise. Find and discuss all of the global and local extreme points for the following functions. Say whether the extreme points are maxima or minima, and whether they are global or local.

a) $f(x) = x^4 - x^2$ for $-2 \le x \le 2$.

b) $g(x) = 4x^3 - 3x^4$ for $-2 \le x \le 2$.

12.4 The Mean Value Theorem

12.24 Lemma (Rolle's Theorem) *Let a, b be real numbers with $a < b$ and let $f\colon [a, b] \to \mathbf{R}$ be a function that is continuous on $[a, b]$ and differentiable on (a, b). Suppose that $f(a) = f(b)$. Then there is a point $c \in (a, b)$ such that $f'(c) = 0$.*

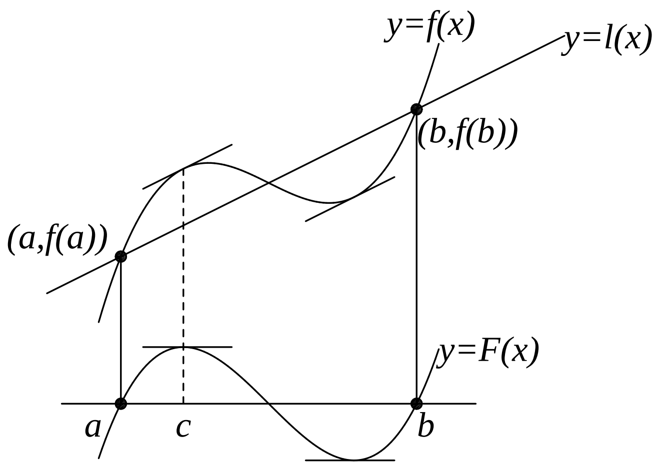
Proof: By the extreme value property, f has a maximum at some point $A \in [a, b]$. If $A \in (a, b)$, then $f'(A) = 0$ by the critical point theorem. Suppose $A \in \{a, b\}$. By the extreme value property, f has a minimum at some point $B \in [a, b]$. If $B \in (a, b)$ then $f'(B) = 0$ by the critical point theorem. If $B \in \{a, b\}$, then we have $\{A, B\} \subset \{a, b\}$ so $f(A) = f(B) = f(a) = f(b)$. Hence in this case the maximum value and the minimum value taken by f are equal, so $f(x) = f(a)$ for $x \in [a, b]$ so $f'(x) = 0$ for all $x \in (a, b)$. |||

Rolle's theorem is named after Michel Rolle (1652-1719). An English translation of Rolle's original statement and proof of the theorem can be found in [43, pages 253-260]. It takes a considerable effort to see any relation between what Rolle says and what our form of Rolle's theorem says.

12.25 Theorem (Mean value theorem.) *Let a, b be real numbers and let $f\colon [a, b] \to \mathbf{R}$ be a function that is continuous on $[a, b]$ and differentiable on (a, b). Then there is a point $c \in (a, b)$ such that $f'(c) = \dfrac{f(b) - f(a)}{b - a}$; i.e., there is a point c where the slope of the tangent line is equal to the slope of the line joining $\big(a, f(a)\big)$ to $\big(b, f(b)\big)$.*

Proof: The equation of the line joining $\big(a, f(a)\big)$ to $\big(b, f(a)\big)$ is

$$y = l(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$



Let

$$\begin{aligned} F(x) &= f(x) - l(x) \\ &= f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a). \end{aligned}$$

Then F is continuous on $[a, b]$ and differentiable on (a, b) and $F(a) = F(b) = 0$. By Rolle's theorem there is a point $c \in (a, b)$ where $F'(c) = 0$.

Now

$$F'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

so

$$F'(c) = 0 \implies f'(c) - \frac{f(b) - f(a)}{b - a} = 0$$
$$\implies f'(c) = \frac{f(b) - f(a)}{b - a}. \, \|$$

12.26 Corollary. *Let J be an interval in* $\mathbf{R}$ *and let $f\colon J \to \mathbf{R}$ be a function that is continuous on J and differentiable at the interior points of J. Then*

$$f'(x) = 0 \text{ for all } x \in \text{interior}\,(J) \implies f \text{ is constant on } J.$$
$$f'(x) \leq 0 \text{ for all } x \in \text{interior}\,(J) \implies f \text{ is decreasing on } J.$$
$$f'(x) \geq 0 \text{ for all } x \in \text{interior}\,(J) \implies f \text{ is increasing on } J.$$
$$f'(x) < 0 \text{ for all } x \in \text{interior}\,(J) \implies f \text{ is strictly decreasing on } J.$$
$$f'(x) > 0 \text{ for all } x \in \text{interior}\,(J) \implies f \text{ is strictly increasing on } J.$$

Proof: I will prove the second assertion. Suppose $f'(x) \leq 0$ for all $x \in \text{interior}(J)$. Let s, t be points in J with $s < t$. Then by the mean value theorem

$$f(t) - f(s) = f'(c)(t - s) \text{ for some } c \in (s, t).$$

Since $f'(c) \leq 0$ and $(t - s) > 0$, we have $f(t) - f(s) = f'(c)(t - s) \leq 0$; i.e., $f(t) \leq f(s)$. Thus f is decreasing on J. $\|$

12.27 Exercise. Prove the first assertion of the previous corollary; i.e., prove that if f is continuous on an interval J, and $f'(x) = 0$ for all $x \in \text{interior}(J)$, then f is constant on J.

12.28 Definition (Antiderivative) Let f be a real valued function with $\text{dom}(f) \subset \mathbf{R}$. Let J be an interval such that $J \subset \text{dom}(f)$. A function F is an *antiderivative for* f on J if F is continuous on J and $F'(x) = f(x)$ for all x in the interior of J.

12.29 Examples. Since $\dfrac{d}{dx}(x^3 + 4) = 3x^2$, we see that $x^3 + 4$ is an antiderivative for $3x^2$. Since

$$\frac{d}{dx}(\cos^2(x)) = 2\cos(x)(-\sin(x)) = -2\sin(x)\cos(x),$$

and

$$\frac{d}{dx}(-\sin^2(x)) = -2 \cdot \sin(x)\cos(x)$$

we see that $\cos^2$ and $-\sin^2$ are both antiderivatives for $-2\sin \cdot \cos$.

We will consider the problem of finding antiderivatives in chapter 17. Now I just want to make the following observation:

12.30 Theorem (Antiderivative theorem.) *Let f be a real valued function with* $\text{dom}(f) \subset \mathbf{R}$ *and let J be an interval with* $J \subset \text{dom}(f)$. *If F and G are two antiderivatives for f on J, then there is a number $c \in \mathbf{R}$ such that*

$$F(x) = G(x) + c \textit{ for all } x \in J.$$

12.31 Exercise. Prove the antiderivative theorem.

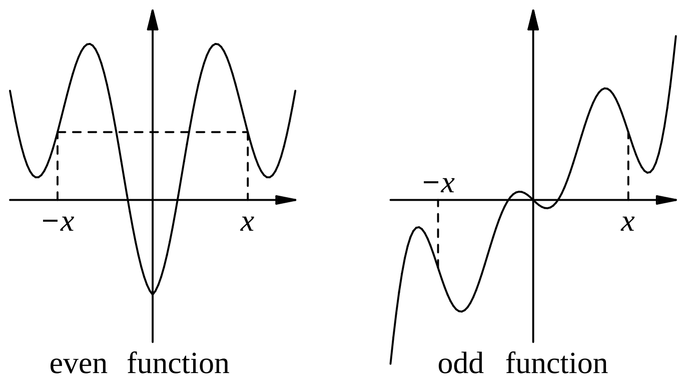
12.32 Definition (Even and odd functions.) A subset S of $\mathbf{R}$ is called *symmetric* if ($x \in S \implies -x \in S$). A function f is said to be *even* if $\text{dom}(f)$ is a symmetric subset of $\mathbf{R}$ and

$$f(x) = f(-x) \text{ for all } x \in \text{dom}(f),$$

and f is said to be *odd* if $\text{dom}(f)$ is a symmetric subset of $\mathbf{R}$ and

$$f(x) = -f(-x) \text{ for all } x \in \text{dom}(f)$$

.



even function odd function

12.33 Example. If $n \in \mathbf{Z}^+$ and $f(x) = x^n$, then f is even if n is even, and f is odd if n is odd. Also cos is an even function and sin is an odd function, while ln is neither even or odd.

12.34 Example. If f is even, then $V\Big(\text{graph}(f)\Big) = \text{graph}(f)$ where V is the reflection about the vertical axis. If f is odd, then $R_\pi\Big(\text{graph}(f)\Big) = \text{graph}(f)$ where R_π is a rotation by π about the origin.

12.35 Exercise. Are there any functions that are both even and odd?

12.36 Exercise.

a) If f is an arbitrary even differentiable function, show that the derivative of f is odd.

b) If g is an arbitrary odd differentiable function, show that the derivative of g is even.

Chapter 13

Applications

13.1 Curve Sketching

13.1 Example. Let $f(x) = \dfrac{x^3}{1-x^2}$. Here $\mathrm{dom}(f) = \mathbf{R} \setminus \{\pm 1\}$ and f is an odd function. We have

$$f'(x) = \frac{(1-x^2)3x^2 - x^3(-2x)}{(1-x^2)^2} = \frac{3x^2 - x^4}{(1-x^2)^2} = \frac{x^2(3-x^2)}{(1-x^2)^2}.$$

From this we see that the critical set for f is $\{0, \sqrt{3}, -\sqrt{3}\}$. We can determine the sign of $f'(x)$ by looking at the signs of its factors: Since f is odd, I will consider only points where $x > 0$.

	$0 < x < 1$	$1 < x < \sqrt{3}$	$\sqrt{3} < x$
$\frac{x^2}{(1-x^2)^2}$	$+$	$+$	$+$
$3 - x^2$	$+$	$+$	$-$
$f'(x)$	$+$	$+$	$-$

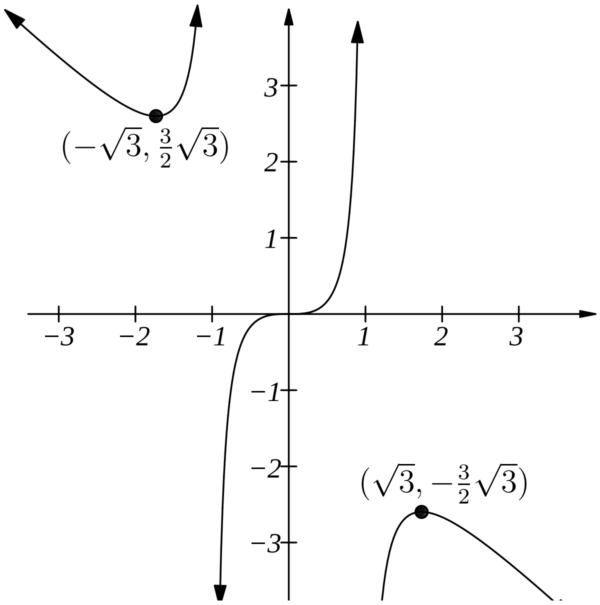
Thus f is strictly increasing on $(0, 1)$ and on $(1, \sqrt{3})$, and f is strictly decreasing on $(\sqrt{3}, \infty)$. Also

$$f(\sqrt{3}) = \frac{3\sqrt{3}}{1-3} = -\frac{3}{2}\sqrt{3} \text{ and } f(0) = 0.$$

We see that $|f(x)|$ is unbounded on any interval $(1-\delta, 1)$ or $(1, 1+\delta)$, since the numerator of the fraction is near to 1, and the denominator is near to 0 on these intervals. Also

$$f(x) = \frac{x^3}{1-x^2} = x\Big(\frac{x^2}{1-x^2}\Big) = x\Big(\frac{1}{-1+\frac{1}{x^2}}\Big),$$

so $|f(x)|$ is large when x is large. ($f(x)$ is the product of x and a number near to -1.) Using this information we can make a reasonable sketch of the graph of f.



Here f has a local maximum at $\sqrt{3}$ and a local minimum at $-\sqrt{3}$. It has no global extreme points.

13.2 Definition (Infinite limits.) Let $\{x_n\}$ be a real sequence. We say

$$\lim\{x_n\} = +\infty \text{ or } \{x_n\} \to +\infty$$

if for every $B \in \mathbf{R}$ there is an $N \in \mathbf{Z}^+$ such that for all $n \in \mathbf{Z}_{\geq N}$ $\Big(x_n > B\Big)$.

We say

$$\lim\{x_n\} = -\infty \text{ or } \{x_n\} \to -\infty$$

if for every $B \in \mathbf{R}$ there is an $N \in \mathbf{Z}^+$ such that for all $n \in \mathbf{Z}_{\geq N}$ $\Big(x_n < B\Big)$.
Let f be a real valued function such that $\mathrm{dom}(f) \subset \mathbf{R}$, and let $a \in \mathbf{R}$. We say

$$\lim_{x \to a^+} f(x) = +\infty$$

if $\mathrm{dom}(f)$ contains an interval $(a, a+\epsilon)$ and for every sequence $\{x_n\}$ in $\mathrm{dom}(f) \cap (a, \infty)$

$$(\{x_n\} \to a) \implies \Big(\{f(x_n)\} \to +\infty\Big).$$

We say

$$\lim_{x \to a^-} f(x) = +\infty$$

if $\mathrm{dom}(f)$ contains an interval $(a-\epsilon, a)$ and for every sequence $\{x_n\}$ in $\mathrm{dom}(f) \cap (-\infty, a)$

$$(\{x_n\} \to a) \implies \Big(\{f(x_n)\} \to +\infty\Big).$$

Similar definitions can be made for

$$\lim_{x \to a^+} f(x) = -\infty, \quad \lim_{x \to a^-} f(x) = -\infty.$$

We say $\lim_{x \to +\infty} f(x) = +\infty$ if $\mathrm{dom}(f)$ contains some interval (a, ∞) and for every sequence $\{x_n\}$ in $\mathrm{dom}(f)$

$$\{x_n\} \to +\infty \implies \{f(x_n)\} \to +\infty.$$

Similarly if $c \in \mathbf{R}$ we can define

$$\lim_{x \to +\infty} f(x) = -\infty, \quad \lim_{x \to +\infty} f(x) = c, \quad \lim_{x \to -\infty} f(x) = +\infty, \text{ etc.}$$

13.3 Example. If f is the function in the previous example (i.e. $f(x) = \dfrac{x^3}{1-x^2}$) then

$$\begin{aligned} \lim_{x \to 1^+} f(x) &= -\infty, \\ \lim_{x \to 1^-} f(x) &= +\infty, \\ \lim_{x \to +\infty} f(x) &= -\infty, \end{aligned}$$

and

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

Also,

$$\begin{aligned}\lim_{x\to\infty}\frac{1}{x} &= 0,\\ \lim_{x\to 0^+}\frac{x}{|x|} &= 1,\\ \lim_{x\to 0^-}\frac{x}{|x|} &= -1,\end{aligned}$$

and

$$\lim_{x\to+\infty}\frac{x^2+1}{x^2+3x}=\lim_{x\to+\infty}\frac{1+\frac{1}{x^2}}{1+\frac{3}{x}}=1.$$

The situation here is very similar to the situation in the case of ordinary limits, and we will proceed without writing out detailed justifications.

13.4 Exercise. Write out definitions for

$$\Big(\lim_{x\to+\infty}f(x)=-\infty\Big)\text{ and for }\Big(\lim_{x\to a^-}f(x)=-\infty\Big).$$

13.5 Exercise. Find one function f satisfying all of the following conditions:

$$\begin{aligned}\lim_{x\to+\infty}f(x) &= 3,\\ \lim_{x\to 3^+}f(x) &= +\infty,\\ \lim_{x\to 3^-}f(x) &= +\infty.\end{aligned}$$

13.6 Example. Let $f(x)=\sin(2x)+2\sin(x)$. Then $f(x+2\pi)=f(x)$ for all $x\in\mathbf{R}$, so I will restrict my attention to the interval $[-\pi,\pi]$. Also f is an odd function, so I will further restrict my attention to the interval $[0,\pi]$. Now

$$\begin{aligned}f'(x) &= 2\cos 2x+2\cos x=2(2\cos^2 x-1)+2\cos x\\ &= 2(2\cos^2 x+\cos x-1)=2(2\cos x-1)(\cos x+1)\\ &= 4\Big(\cos x-\frac{1}{2}\Big)(\cos x+1).\end{aligned}$$

Hence x is a critical point for f if and only if $\cos x\in\Big\{\frac{1}{2},-1\Big\}$. The critical points of f in $[0,\pi]$ are thus π and $\frac{\pi}{3}$, and the critical points in $[-\pi,\pi]$ are

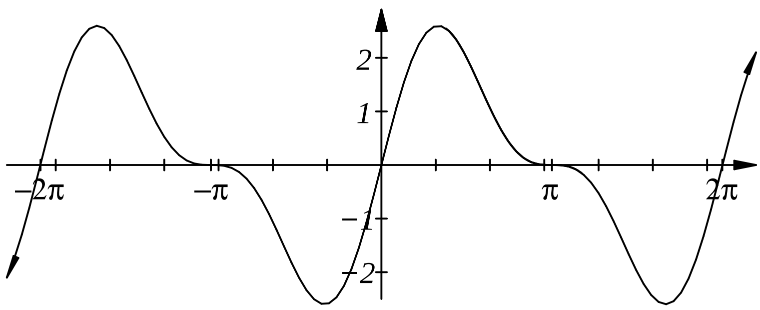
$\left\{-\pi, \pi, \frac{\pi}{3}, -\frac{\pi}{3}\right\}$. Now $f(\pi) = f(0) = 0$ and

$$f\Big(\frac{\pi}{3}\Big) = \sin\Big(\frac{2\pi}{3}\Big) + 2\sin\Big(\frac{\pi}{3}\Big) = \frac{\sqrt{3}}{2} + \frac{2\sqrt{3}}{2} = \frac{3\sqrt{3}}{2} = 2.6\text{(approximately)},$$

and $f\Big(-\frac{\pi}{3}\Big) = -f\Big(\frac{\pi}{3}\Big)$. Also note $f'(0) = 4$. Since f is continuous on $[-\pi, \pi]$, we know that f has a maximum and a minimum on this interval, and since $f(x + 2\pi) = f(x)$ for all $x \in \mathbf{R}$, the maximum (or minimum) of f on $[-\pi, \pi]$ will be a global maximum (or minimum) for f. Since f is differentiable everywhere, the extreme points are critical points and from our calculations f has a maximum at $\frac{\pi}{3}$ and a minimum at $-\frac{\pi}{3}$. I will now determine the sign of f' on $[0, \pi]$:

	$0 < x < \frac{\pi}{3}$	$\frac{\pi}{3} < x < \pi$
$\cos x + 1$	$+$	$+$
$\cos x - \frac{1}{2}$	$+$	$-$
$f'(x)$	$+$	$-$

Thus f is strictly increasing on $\Big(0, \frac{\pi}{3}\Big)$ and f is strictly decreasing on $\Big(\frac{\pi}{3}, \pi\Big)$. We can now make a reasonable sketch for the graph of f.



13.7 Exercise. Sketch and discuss the graphs of the following functions. Mention all critical points and determine whether each critical point is a local or global maximum or minimum.

a) $f(x) = (1 - x^2)^2$.

b) $g(x) = \dfrac{x}{1+x^2}$.

c) $h(x) = x + \sin(x)$.

d) $k(x) = x\ln(x)$.

(The following remark may be helpful for determining $\lim\limits_{x\to 0} k(x)$. If $0 < t < 1$, then $\dfrac{1}{t} < \dfrac{1}{t^{\frac{3}{2}}}$. Hence if $0 < x < 1$, then

$$\begin{aligned} |\ln(x)| &= \left|\int_1^x \frac{1}{t}dt\right| = \int_x^1 \frac{1}{t}dt \le \int_x^1 \frac{1}{t^{\frac{3}{2}}}dt \\ &= -\frac{2}{t^{\frac{1}{2}}}\Big|_x^1 = 2\left(\frac{1}{\sqrt{x}} - 1\right) \le \frac{2}{\sqrt{x}}. \end{aligned}$$

Thus,

$$|x\ln(x)| \le 2\sqrt{x} \text{ for } 0 < x < 1).$$

13.2 Optimization Problems.

13.8 Example. A stick of length l is to be broken into four pieces of length s, s, t and t and the pieces are to be assembled to make a rectangle. How should s and t be chosen if the area of the rectangle is to be as large as possible? What is the area of this largest rectangle? Before doing the problem you should guess the answer. Your guess will probably be correct.

Let s be the length of one side of the rectangle. Then $2s + 2t = l$ so $t = \dfrac{l}{2} - s$; i.e., t is a function of s. Let $A(s)$ be the area of a rectangle with side s. Then

$$A(s) = st = s\Big(\frac{l}{2} - s\Big) = \frac{l}{2}s - s^2 \quad \text{for } 0 \le s \le \frac{l}{2}.$$

I include the endpoints for convenience; i.e., I consider rectangles with zero area to be admissible candidates for my answer. These clearly correspond to minimum area. Now

$$A'(s) = \frac{l}{2} - 2s$$

so A has only one critical point, namely $\frac{l}{4}$, and

$$A\Big(\frac{l}{4}\Big) = \frac{l}{2}\frac{l}{4} - \Big(\frac{l}{4}\Big)^2 = \frac{l^2}{16} = \Big(\frac{l}{4}\Big)^2.$$

Since A is continuous on $\Big[0, \frac{l}{2}\Big]$ we know that A has a maximum and a minimum, and since A is differentiable on $\Big(0, \frac{l}{2}\Big)$ the extreme points are a subset of $\Big\{0, \frac{l}{2}, \frac{l}{4}\Big\}$. Since $A(0) = A\Big(\frac{l}{2}\Big) = 0$ the maximal area is $\Big(\frac{l}{4}\Big)^2$; i.e., the maximal rectangle is a square. (As you probably guessed.)

This problem is solved by Euclid in completely geometrical terms [17, vol 1 page 382].

Euclid's proof when transformed from geometry to algebra becomes the following. Suppose in our problem $s \neq t$, say $s < t$. Since $s + t = \frac{l}{2}$, it follows that $s \leq \frac{l}{4} \leq t$ (if s and t were both less than $\frac{l}{4}$, we'd get a contradiction, and if they were both greater than $\frac{l}{4}$, we'd get a contradiction). Let r be defined by

$$s = \frac{l}{4} - r \text{ so } r \geq 0.$$

Then $t = \frac{l}{2} - s = \frac{l}{2} - \Big(\frac{l}{4} - r\Big) = \frac{l}{4} + r$ so

$$A(s) = st = \Big(\frac{l}{4} - r\Big)\Big(\frac{l}{4} + r\Big) = \Big(\frac{l}{4}\Big)^2 - r^2 = A\Big(\frac{l}{4}\Big) - r^2.$$

Hence, if $r > 0$, $A(s) < A\Big(\frac{l}{4}\Big)$ and to get a maximum we must have $r = 0$ and $s = \frac{l}{4}$. This proof requires knowing the answer ahead of time (but you probably were able to guess it). In any case, Euclid's argument is special, whereas our calculus proof applies in many situations.

Quadratic polynomials can be minimized (or maximized) without calculus by completing the square. For example, we have

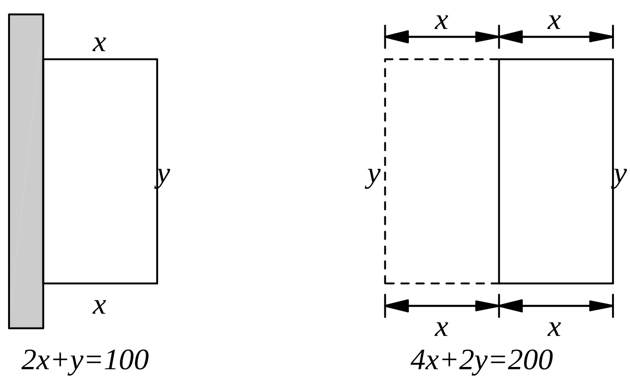
$$A(s) \quad = \quad -s^2 + \frac{l}{2}s$$

$$= -\left(s^2 - \frac{l}{2}s + \left(\frac{l}{4}\right)^2\right) + \left(\frac{l}{4}\right)^2$$
$$= \left(\frac{l}{4}\right)^2 - \left(s - \frac{l}{4}\right)^2.$$

From this we can easily see that $A(s) \leq \left(\frac{l}{4}\right)^2$ for all s and equality holds only if $s = \frac{l}{4}$. This technique applies only to quadratic polynomials.

13.9 Example. Suppose I have 100 ft. of fence, and I want to fence off 3 sides of a rectangular garden, the fourth side of which lies against a wall and requires no fence (see the figure). What should the sides of the garden be if the area is to be as large as possible?

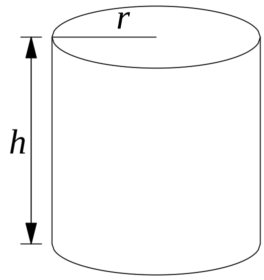
This is a straightforward problem, and in the next exercise you will do it by using calculus. Here I want to indicate how to do the problem without calculation. Imagine that the wall is a mirror, and that my fence is reflected in the wall.



When I maximize the area of a garden with a rectangle of sides x and y, then I have maximized the area of a rectangle bounded by 200 feet of fence (on four sides) with sides y and $2x$. From the previous problem the answer to this problem is a square with $y = 2x = 50$. Hence, the answer to my original question is $y = 50$, $x = 25$. Often optimization problems have solutions that can be guessed on the basis of symmetry. You should try to guess answers to these problems before doing the calculations.

13.10 Exercise. Verify my solution in the previous example by using calculus *and* by completing the square.

13.11 Example. I want to design a cylindrical can of radius r and height h with a volume of V_0 cubic feet (V_0 is a constant). How should I choose r and h if the amount of tin in the can is to be minimum?



Here I don't see any obvious guess to make for the answer.

I have

$$V_0 = \text{ volume of can } = \pi r^2 h,$$

so $h = \dfrac{V_0}{\pi r^2}$. Let $A(r)$ be the surface area of the can of radius r. Then

$$\begin{aligned} A(r) &= \text{ area of sides } + 2 \text{ (area of top)} \\ &= 2\pi r \cdot h + 2(\pi r^2) \\ &= 2\pi r \frac{V_0}{\pi r^2} + 2\pi r^2 \\ &= \frac{2V_0}{r} + 2\pi r^2. \end{aligned}$$

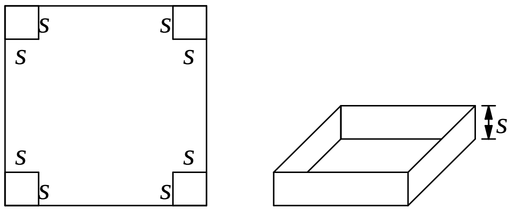
The domain of A is $\mathbf{R}^+$. It is clear that $\lim_{r \to 0^+} A(r) = +\infty$, and $\lim_{r \to +\infty} A(r) = +\infty$. Now $A'(r) = -\dfrac{2V_0}{r^2} + 4\pi r = \dfrac{4\pi}{r^2}\Big(r^3 - \dfrac{V_0}{2\pi}\Big)$. The only critical point for A is $r = \sqrt[3]{\dfrac{V_0}{2\pi}}$ (call this number r_0). Then $A'(r) = \dfrac{4\pi}{r^2}(r^3 - r_0^3)$. We see that $A'(r) < 0$ for $r \in (0, r_0)$ and $A'(r) > 0$ for $r \in (r_0, \infty)$ so A is decreasing on $(0, r_0]$ and A is increasing on $[r_0, \infty)$ and thus A has a minimum at r_0. The value of h corresponding to r_0 is

$$h = \frac{V_0}{\pi r_0^2} = \frac{V_0}{\pi\Big(\frac{V_0}{2\pi}\Big)^{2/3}} = \frac{2^{2/3}}{\pi^{1/3}} V_0^{1/3} = 2\Big(\frac{V_0^{1/3}}{2^{1/3}\pi^{1/3}}\Big) = 2r_0.$$

Thus the height of my can is equal to its diameter; i.e., the can will exactly fit into a cubical box.

In the following four exercises see if you can make a reasonable guess to the solutions before you use calculus to find them.

13.12 Exercise. A box (without a lid) is to be made by cutting 4 squares of side s from the corners of a $12'' \times 12''$ square, and folding up the corners as indicated in the figure.

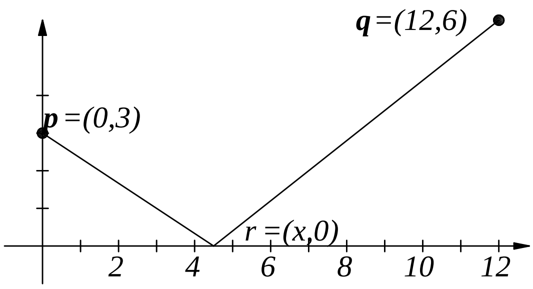


How should s be chosen to make the volume of the box as large as possible?

13.13 Exercise. A rectangular box with a square bottom and no lid is to be built having a volume of 256 cubic inches. What should the dimensions be, if the total surface area of the box is to be as small as possible?

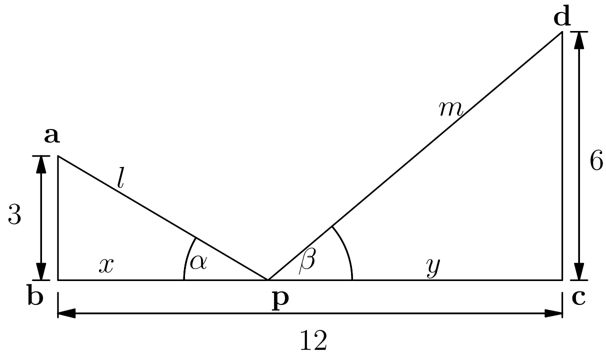
13.14 Exercise. Find the point(s) on the parabola whose equation is $y = x^2$ that are nearest to the point $\left(0, \frac{9}{2}\right)$.

13.15 Exercise. Let $\mathbf{p} = (0, 3)$ and let $\mathbf{q} = (12, 6)$. Find the point(s) $\mathbf{r}$ on the x-axis so that path from $\mathbf{p}$ to $\mathbf{r}$ to $\mathbf{q}$ is as short as possible; i.e., such that length([$\mathbf{pr}$]) + length([$\mathbf{rq}$]) is as short as possible.



You don't need to prove that the critical point(s) you find are actually minimum points.

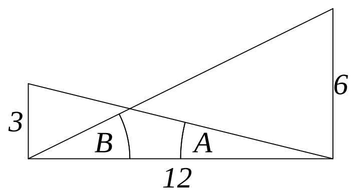
13.3 Rates of Change



Suppose in the given figure, I want to find the shortest path from **a** to a point **p** on the segment [**b c**] and back to **d**. Any such path will be uniquely defined by giving any one of the six numbers:

$$\begin{array}{ll} x = \text{dist}(\mathbf{b},\mathbf{p}), & 0 \le x \le 12. \\ y = \text{dist}(\mathbf{p},\mathbf{c}), & 0 \le y \le 12 \\ l = \text{dist}(\mathbf{a},\mathbf{p}), & 3 \le l \le \sqrt{3^2+12^2}. \\ m = \text{dist}(\mathbf{d},\mathbf{p}), & 6 \le m \le \sqrt{6^2+12^2}. \\ \alpha = \angle\mathbf{apb}, & A \le \alpha \le \frac{\pi}{2}. \\ \beta = \angle\mathbf{dpc}, & B \le \beta \le \frac{\pi}{2}. \end{array}$$

Here A, B are as shown in the figure below:



$$\frac{3}{12} = \frac{\sin(A)}{\cos(A)} = \tan(A) \text{ and } \frac{6}{12} = \frac{\sin(B)}{\cos(B)} = \tan(B).$$

For a given point **p**, any of the six numbers is a function of any of the others. For example, we have l is a function of x

$$l = \sqrt{x^2+9},$$

and l is a function of m, since for $x \in [0, 12]$ and $y \in [0, 12]$ we have

$$\begin{aligned} m^2 = (12-x)^2 + 36 &\implies 12 - x = \sqrt{m^2 - 36} \\ &\implies 12 - \sqrt{m^2 - 36} = x \\ &\implies l = \sqrt{\left(12 - \sqrt{m^2-36}\right)^2 + 9}. \end{aligned}$$

Also l is a function of α, since by similar triangles $\dfrac{\sin(\alpha)}{1} = \dfrac{3}{l}$ and hence

$$l = \frac{3}{\sin(\alpha)} = 3\csc(\alpha).$$

We have

$$\frac{dl}{dx} = \frac{1}{2}\frac{2x}{\sqrt{x^2+9}} = \frac{x}{\sqrt{x^2+9}}$$

and

$$\frac{dl}{d\alpha} = -3\csc(\alpha)\cot(\alpha).$$

I refer to $\dfrac{dl}{dx}$ as the *rate of change of l with respect to x* and to $\dfrac{dl}{d\alpha}$ as the *rate of change of l with respect to α*. Note that the "l"'s in "$\dfrac{dl}{dx}$"and "$\dfrac{dl}{d\alpha}$"represent different functions. In the first case $l(x) = \sqrt{x^2+9}$ and in the second case $l(\alpha) = 3\csc\alpha$. Here $\dfrac{dl}{dx}$ is positive, indicating that l increases when x increases, and $\dfrac{dl}{d\alpha}$ is negative, indicating that l decreases when α increases.

I want to find the path for which $l + m$ is shortest; i.e., I want to find the minimum value of $l + m$. I can think of l and m as being functions of x, and then the minimum value will occur when $\dfrac{d}{dx}(l+m) = 0$; i.e.,

$$\frac{dl}{dx} + \frac{dm}{dx} = 0. \tag{13.16}$$

Now $l^2 = x^2 + 9$, so $2l \cdot \dfrac{dl}{dx} = 2 \cdot x$; i.e.,

$$\frac{dl}{dx} = \frac{x}{l} = \cos\alpha,$$

and $m^2 = (12-x)^2 + 6^2$, so $2m\dfrac{dm}{dx} = 2(12-x)(-1)$, i.e.,

$$\frac{dm}{dx} = -\frac{(12-x)}{m} = -\frac{y}{m} = -\cos\beta.$$

Equation (13.16) thus says that for the minimum path $\cos\alpha - \cos\beta = 0$; i.e., $\cos\alpha = \cos\beta$, and hence $\alpha = \beta$. Thus the minimizing path satisfies the reflection condition, *angle of incidence equals angle of reflection.* Hence the minimizing triangle will make $\triangle\mathbf{bpa}$ and $\triangle\mathbf{cpd}$ similar, and will satisfy

$$\frac{6}{y} = \frac{3}{x} \text{ and } x + y = 12,$$

so

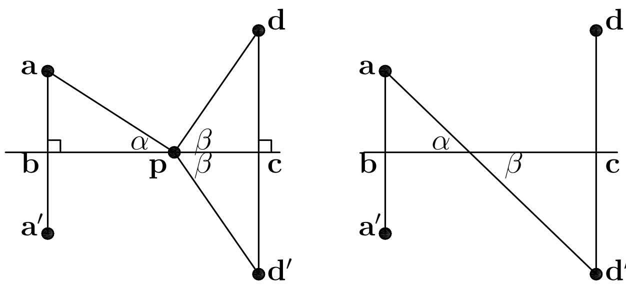
$$6x = 3y = 3(12-x) = 36 - 3x$$

or

$$9x = 36 \text{ so } x = 4 \text{ and } y = 8.$$

This example was done pretty much as Leibniz would have done it. You should compare the solution given here to your solution of exercise 13.15.

The problem in the last example was solved by Heron (date uncertain, sometime between 250 BC and 150 AD) as follows[26, page 353]. Imagine the line $[\mathbf{bc}]$ to be a mirror. Let $\mathbf{a}'$ and $\mathbf{d}'$ denote the images of $\mathbf{a}$ and $\mathbf{d}$ in the mirror,



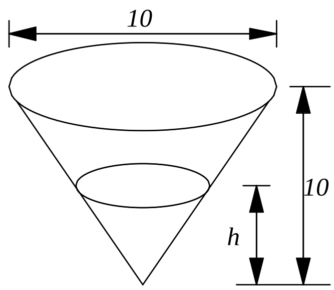
i.e. $[\mathbf{aa}']$ and $[\mathbf{dd}']$ are perpendicular to $[\mathbf{bc}]$ and $\text{dist}(\mathbf{a},\mathbf{b}) = \text{dist}(\mathbf{a}',\mathbf{b})$, $\text{dist}(\mathbf{d},\mathbf{c}) = \text{dist}(\mathbf{d}',\mathbf{c})$. Consider any path $\mathbf{apd}$ going from $\mathbf{a}$ to a point $\mathbf{p}$ on the mirror, and then to $\mathbf{d}$. Then triangle($\mathbf{pcd}$) and triangle($\mathbf{pcd}'$) are congruent, and hence

$$\text{dist}(\mathbf{p},\mathbf{d}) = \text{dist}(\mathbf{p},\mathbf{d}').$$

and hence the paths **apd** and **apd**$'$ have equal lengths. Now the shortest path **apd**$'$ is a straight line, which makes the angles α and β are vertical angles, which are equal. Hence the shortest path makes the angle of incidence equal to the angle of reflection, as we found above by calculus.

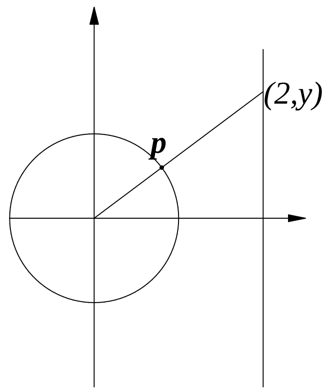
Remark: We can think of velocity as being rate of change of position with respect to time.

13.17 Exercise. Consider a conical tank in the shape of a right circular cone with altitude $10'$ and diameter $10'$ as shown in the figure.



Water flows into the tank at a constant rate of 10 cubic ft./minute. Let h denote the height of the water in the tank at a given time t. Find the rate of change of h with respect to t. What is this rate when the height of the water is $5'$? What can you say about $\dfrac{dh}{dt}$ when h is nearly zero?

13.18 Exercise. A particle **p** moves on the rim of a wheel of radius 1 that rotates about the origin at constant angular speed ω, so that at time t it is at the point $\big(\cos(\omega t), \sin(\omega t)\big)$. A light at the origin causes **p** to cast a shadow at the point $(2, y)$ on a wall two feet from the center of the wheel.



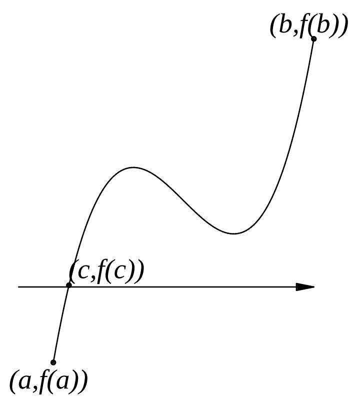
Find the rate of change of y with respect to time. You should ignore the speed of light, i.e. ignore the time it takes light to travel from the origin to the wall.

Chapter 14

The Inverse Function Theorem

14.1 The Intermediate Value Property

14.1 Assumption (Intermediate value property 1.) Let a, b be real numbers with $a < b$, and let f be a continuous function from $[a, b]$ to $\mathbf{R}$ such that $f(a) < 0$ and $f(b) > 0$. Then there is some number $c \in (a, b)$ such that $f(c) = 0$.



The intermediate value *theorem* was first proved in 1817 by Bernard Bolzano (1781–1848). However Bolzano published his proof in a rather obscure Bohemian journal, and his work did not become well known until much later. Before the nineteenth century the theorem was often assumed implicitly, i.e. it was used without stating that it was an assumption.

14.2 Definition (c is between a and b.) Let a, b and c be real numbers with $a \neq b$. We say that c is *between* a and b if either $a < c < b$ or $b < c < a$.

14.3 Corollary (Intermediate value property 2.) *Let f be a continuous function from some interval $[a, b]$ to $\mathbf{R}$, such that $f(a)$ and $f(b)$ have opposite signs. Then there is some number c between a and b such that $f(c) = 0$.*

Proof: If $f(a) < 0 < f(b)$ the result follows from assumption 14.1. Suppose that $f(b) < 0 < f(a)$. Let $g(x) = -f(x)$ for all $x \in [a, b]$. then g is a continuous function on $[a, b]$ and $g(a) < 0 < g(b)$. It follows that there is a number $c \in (a, b)$ such that $g(c) = 0$, and then $f(c) = -g(c) = 0$. |||

14.4 Corollary (Intermediate value property 3.) *Let a, b be real numbers with $a < b$, and let $f : [a, b] \to \mathbf{R}$ be a continuous function such that $f(a) \neq f(b)$. Let p be any number between $f(a)$ and $f(b)$. Then there is a number $c \in (a, b)$ such that $f(c) = p$.*

14.5 Exercise. Prove Corollary 14.4. You may assume that $f(a) < f(b)$.

14.2 Applications

14.6 Example. We know that ln is continuous on $\mathbf{R}^+$, and that $\ln(2) \leq 1 \leq \ln(4)$.(Cf equation (5.78).) It follows that there is a number e in $[2, 4]$ such that $\ln(e) = 1$.

14.7 Example. Two points P, Q on a sphere are called *antipodal points* if P and Q are opposite ends of the same diameter of the sphere. We will consider the surface of the earth to be a sphere of radius R. At any fixed time, let $T(p)$ denote the temperature of the earth at the point p on the surface of the earth. (More precisely, let $T(p)$ be the number such that the temperature at p is $T(p)°C$). We will show that there are two antipodal points P, Q on the surface of the earth such that $T(P) = T(Q)$. In fact, we will show that there are two antipodal points on the equator with the same temperature. We first introduce a coordinate system so that the center of the earth is at the origin, and the plane of the equator is the x-y plane, and the point on the equator passing through the Greenwich meridian is the point $(R, 0)$. Then the points on the equator are the points

$$(R\cos(\theta), R\sin(\theta)) \text{ where } \theta \in \mathbf{R}.$$

Define a function $f : [0, \pi] \to \mathbf{R}$ by

$$f(\theta) = T(R\cos(\theta), R\sin(\theta)) - T(-R\cos(\theta), -R\sin(\theta)).$$

Thus

$$f(0) = T(R, 0) - T(-R, 0).$$

We suppose that f is a continuous function on $[0, \pi]$. If $f(0) = 0$ then $T(R, 0) = T(-R, 0)$, so $(R, 0)$ and $(-R, 0)$ are a pair of antipodal points with the same temperature. Now

$$f(\pi) = T(-R, 0) - T(R, 0) = -f(0),$$

so if $f(0) \neq 0$ then $f(0)$ and $f(\pi)$ have opposite signs. Hence by the intermediate value property, there is a number $c \in (0, \pi)$ such that $f(c) = 0$, i.e.

$$T(R\cos(c), R\sin(c)) = T(-R\cos(c), -R\sin(c)).$$

Then $(R\cos(c), R(\sin(c))$ and $(-R\cos(c), -R(\sin(c))$ are a pair of antipodal points with the same temperature. $\|$

14.8 Example. Let

$$P = a_0 + a_1X + a_2X^2 + a_3X^3$$

where a_0, a_1, a_2, and a_3 are real numbers, and $a_3 \neq 0$. Then there exists some number $r \in \mathbf{R}$ such that $P(r) = 0$.
Proof: I will suppose that $P(t) \neq 0$ for all $t \in \mathbf{R}$ and derive a contradiction. Let

$$Q(x) = \frac{P(x)}{P(-x)} \text{ for all } x \in \mathbf{R}.$$

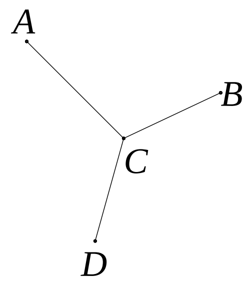
Since $P(x) \neq 0$ for all $x \in \mathbf{R}$, Q is continuous on $\mathbf{R}$. We know that

$$\begin{aligned} \lim\{Q(n)\} &= \lim\{\frac{a_0 + a_1n + a_2n^2 + a_3n^3}{a_0 - a_1n + a_2n^2 - a_3n^3}\} \\ &= \lim\left\{\frac{\frac{a_0}{n^3} + \frac{a_1}{n^2} + \frac{a_2}{n} + a_3}{\frac{a_0}{n^3} - \frac{a_1}{n^2} + \frac{a_2}{n} - a_3}\right\} = -1. \end{aligned}$$

Hence $Q(N) < 0$ for some $N \in \mathbf{Z}^+$. Then $P(N)$ and $P(-N)$ have opposite signs, so by the intermediate value property there is a number $r \in [-N, N]$ such that $P(r) = 0$. This contradicts our assumption that $P(t) \neq 0$ for all $t \in \mathbf{R}$. $\|$

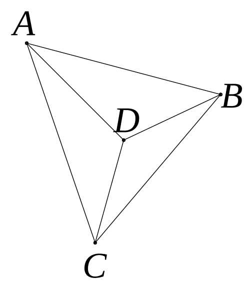
14.9 Exercise. Let $p(x) = x^3 - 3x + 1$. Show that there are at least three different numbers a, b, c such that $p(a) = p(b) = p(c) = 0$.

14.10 Exercise. Three wires AC, BC, DC are joined at a common point C.



Let S be the Y-shaped figure formed by the three wires. Prove that at any time there are two points in S with the same temperature.

14.11 Exercise. Six wires are joined to form the figure F shown in the diagram.



Show that at any time there are three points in F that have the same temperature. To simplify the problem, you may assume that the temperatures at A,B, C, and D are all distinct.

14.3 Inverse Functions

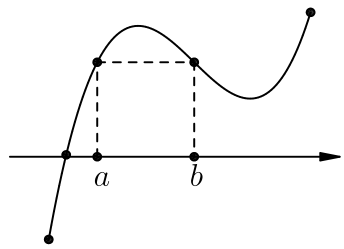
14.12 Definition (Injective.) Let A and B be sets. A function $f : A \to B$ is called *injective* or *one-to-one* if and only if for all points a, b in A

$$(a \neq b) \implies (f(a) \neq f(b)),$$

or equivalently if and only if

$$(f(a) = f(b)) \implies (a = b).$$

If f is a function whose domain and codomain are subsets of $\mathbf{R}$ then f is injective if and only if each horizontal line intersects the graph of f at most once.



$a \neq b,\ f(a) = f(b),$ $\qquad f$ is not injective

14.13 Examples. Let $f : [0,\infty) \to \mathbf{R}$ and $g : \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) = x^2 \text{ for all } x \in [0,\infty)$$

$$g(x) = x^2 \text{ for all } x \in (-\infty,\infty).$$

Then f is injective, since for all $x, y \in [0,\infty)$ we have $x + y > 0$, and hence

$$\Big(x^2 = y^2\Big) \implies \Big(x^2 - y^2 = 0\Big) \implies \Big((x-y)(x+y) = 0\Big) \implies \Big(x = y\Big).$$

However g is not injective, since $g(-1) = g(1)$.

14.14 Remark (Strictly monotonic functions are injective.) If h is strictly increasing on an interval J, then h is injective on J, since for all $x, y \in J$

$$\begin{aligned} x \neq y &\implies ((x < y) \text{ or } (y < x)) \\ &\implies ((h(x) < h(y)) \text{ or } ((h(y) < h(x)) \\ &\implies h(x) \neq h(y). \end{aligned}$$

Similarly, any strictly decreasing function on J is injective.

14.15 Definition (Surjective.) Let A, B be sets and let $f : A \to B$. We say that f is *surjective* if and only if $B = \text{image}(f)$, i.e. if and only if for every $b \in B$ there is at least one element a of A such that $f(a) = b$.

14.16 Examples. Let $f : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R} \to [0, \infty)$ be defined by

$$f(x) = x^2 \text{ for all } x \in (-\infty, \infty)$$

$$g(x) = x^2 \text{ for all } x \in [0, \infty).$$

Then g is surjective, since if $x \in [0, \infty)$, then $x = g(\sqrt{x})$, but f is not surjective, since -1 is not in the image of f.

14.17 Exercise. Give examples of functions with the following properties, or else show that no such functions exist.
$f : \mathbf{R} \to \mathbf{R}$, f is injective and surjective.
$g : \mathbf{R} \to \mathbf{R}$, g is injective but not surjective.
$h : \mathbf{R} \to \mathbf{R}$, h is surjective but not injective.
$k : \mathbf{R} \to \mathbf{R}$, k is neither injective nor surjective.

14.18 Definition (Bijective.) Let A, B be sets. A function $f : A \to B$ is called *bijective* if and only if f is both injective and surjective.

14.19 Examples. If $f : [0, \infty) \to [0, \infty)$ is defined by

$$f(x) = x^2 \text{ for all } x \in [0, \infty),$$

then f is bijective.

The function ln is a bijective function from $\mathbf{R}^+$ to $\mathbf{R}$. We know that ln is strictly increasing, and hence is injective. If y is any real number we know that ln takes on values greater than y, and values less that y, so by the intermediate value property (here we use the fact that ln is continuous) it also takes on the value y, i.e. ln is surjective.

14.20 Remark. Let A and B be sets, and let $f : A \to B$ be a bijective function. Let b be a generic element of B. Since f is surjective, there is an element a in A such that $f(a) = b$. Since f is injective this element a is unique, i.e. if a and c are elements of A then

$$\Big(f(a) = b \text{ and } f(c) = b\Big) \Longrightarrow \Big(f(a) = f(c)\Big) \Longrightarrow \Big(a = c\Big).$$

Hence we can define a function $g : B \to A$ by the rule

$$g(b) = \text{ the unique element } a \in A \text{ such that } f(a) = b.$$

Then by definition

$$f(g(b)) = b \text{ for all } b \in B.$$

Now let $a \in A$, so that $f(a) \in B$. It is clear that the unique element s in A such that $f(s) = f(a)$ is $s = a$, and hence

$$g(f(a)) = a \text{ for all } a \in A.$$

14.21 Definition (Inverse function.) Let A, B be sets, and let $f : A \to B$. An *inverse function* for f is a function $g : B \to A$ such that

$$\Big(f(g(b)) = b \text{ for all } b \in B\Big) \text{ and } \Big(g(f(a)) = a \text{ for all } a \in A\Big).$$

14.22 Remark (Bijective functions have inverses.) Notice that in the definition of inverse functions, both the domain and the codomain of f enter in a crucial way. It is clear that if g is an inverse function for f, then f is an inverse function for g. Remark 14.20 shows that every bijective function $f : A \to B$ has an inverse.

14.23 Example. Let $f : [0, \infty)$ be defined by

$$f(x) = x^2 \text{ for all } x \in [0, \infty).$$

We saw above that f is bijective, and hence has an inverse. If

$$g(x) = \sqrt{x} \text{ for all } x \in [0, \infty)$$

Then it is clear that g is an inverse function for f.

We also saw that $\ln : \mathbf{R}^+ \to \mathbf{R}$ is bijective, and so it has an inverse. This inverse is not expressible in terms of any functions we have discussed. We will give it a name.

14.24 Definition ($E(x)$.) Let E denote the inverse of the logarithm function. Thus E is a function from $\mathbf{R}$ to $\mathbf{R}^+$, and it satisfies the conditions

$$\ln(E(x)) = x \text{ for all } x \in \mathbf{R},$$

$$E(\ln(x)) = x \text{ for all } x \in \mathbf{R}^+.$$

We will investigate the properties of E after we have proved a few general properties of inverse functions.

In order to speak of *the inverse* of a function, as we did in the last definition, we should note that inverses are unique.

14.25 Theorem (Uniqueness of inverses.) *Let A, B be sets and let $f : A \to B$. If g and h are inverse functions for f, then $g = h$.*

Proof: If g and h are inverse functions for f then

$$\text{dom}(g) = \text{dom}(h) = \text{ codomain}(f) = B,$$

and

$$\text{codomain}(g) = \text{codomain}(h) = \text{dom}(f) = A.$$

Also for all $x \in B$

$$h(x) = g(f(h(x))) = g(x).$$

(I have used the facts that $y = g(f(y))$ for all $y \in A$, and $f(h(x)) = x$ for all $x \in B$).

14.26 Theorem (Reflection theorem.) *Let $f : A \to B$ be a function which has an inverse function $g : B \to A$. Then for all $(a, b) \in A \times B$*

$$(a, b) \in \text{graph}(f) \iff (b, a) \in \text{graph}(g).$$

Proof: Let $f : A \to B$ be a function that has an inverse function $g : B \to A$. Then for all $(a, b) \in A \times B$

$$\Big(b = f(a)\Big) \implies \Big(g(b) = g(f(a)) = a\Big) \implies \Big(g(b) = a\Big)$$

and

$$\Big(g(b) = a\Big) \implies \Big(b = (f(g(b)) = f(a)\Big) \implies \Big(b = f(a)\Big).$$

Thus

$$\Big(b = f(a)\Big) \iff \Big(a = g(b)\Big).$$

Now

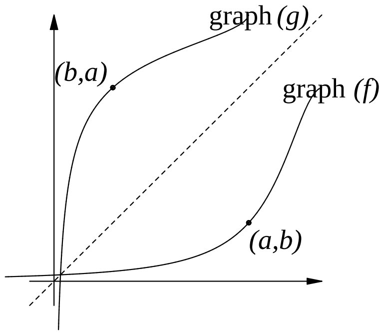
$$\Big(b = f(a)\Big) \iff \Big((a, b) \in \text{ graph}(f)\Big),$$

and

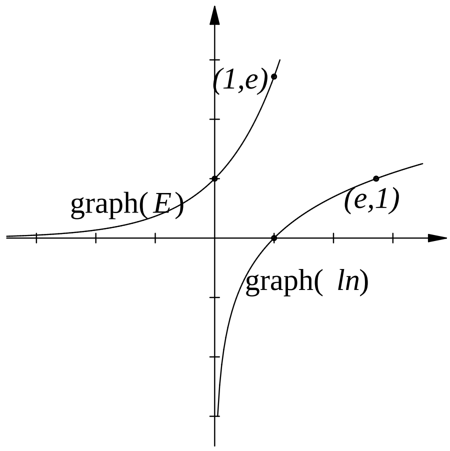
$$\Big(a = g(b)\Big) \iff \Big((b, a) \in \text{ graph}(g)\Big),$$

and the theorem now follows. |||

Remark: If f is a bijective function with $\text{dom}(f) \subset \mathbf{R}$ and $\text{codomain}(f) \subset \mathbf{R}$ Then the reflection theorem says that if g is the inverse function for f, then $\text{graph}(g) = D_+(\text{graph}(f))$ where D_+ is the reflection about the line $y = x$.



Since we know what the graph of ln looks like, we can make a reasonable sketch of graph(E).



It is a standard notation to denote the inverse of a function f by f^{-1}. However since this is also a standard notation for the function $\dfrac{1}{f}$ which is an entirely different object, I will not use this notation.

We have shown that if $f : A \to B$ is bijective, then f has an inverse function. The converse is also true.

14.27 Theorem. *Let A, B be sets and let $f : A \to B$. If f has an inverse function, then f is both injective and surjective.*

Proof: Suppose f has an inverse function $g : B \to A$. Then for all s, t in A we have

$$\Big(f(s) = f(t)\Big) \implies \Big(g(f(s)) = g(f(t))\Big) \implies \Big(s = t\Big) \tag{14.28}$$

and hence f is injective. Also, for each $b \in B$

$$b = f(g(b)),$$

so $b \in \text{image}(f)$, and f is surjective. ‖

14.4 The Exponential Function

14.29 Example. We will now derive some properties of the inverse function E of the logarithm.

We have

$$\begin{aligned} \ln(1) = 0 &\implies E(0) = 1, \\ \ln(e) = 1 &\implies E(1) = e. \end{aligned}$$

For all a and b in $\mathbf{R}$,

$$a + b = \ln(E(a)) + \ln(E(b)) = \ln(E(a)E(b)).$$

If we apply E to both sides of this equality we get

$$E(a + b) = E(a)E(b) \text{ for all } a, b \in \mathbf{R}.$$

For all $a \in \mathbf{R}$ we have

$$1 = E(0) = E(a + (-a)) = E(a)E(-a),$$

from which it follows that

$$E(-a) = (E(a))^{-1} \text{ for all } a \in \mathbf{R}.$$

If $a \in \mathbf{R}$ and $q \in \mathbf{Q}$ we have

$$\ln((E(a))^q) = q \ln(E(a)) = qa.$$

If we apply E to both sides of this identity we get

$$(E(a))^q = E(qa) \text{ for all } a \in \mathbf{R}^+, q \in \mathbf{Q}.$$

In particular,

$$e^q = (E(1))^q = E(q) \text{ for all } q \in \mathbf{Q}. \tag{14.30}$$

Now we have defined $E(x)$ for all $x \in \mathbf{R}$, but we have only defined x^q when $x \in \mathbf{R}^+$ and $q \in \mathbf{Q}$. (We know what $2^{\frac{1}{2}}$ is, but we have not defined $2^{\sqrt{2}}$.) Because of relation (14.30) we often write e^x in place of $E(x)$. E is called the *exponential function*, and is written

$$E(x) = e^x = \exp(x) \text{ for all } x \in \mathbf{R}.$$

We can summarize the results of this example in the following theorem:

14.31 Theorem (Properties of the exponential function.) *The exponential function is a function from* $\mathbf{R}$ *onto* $\mathbf{R}^+$. *We have*

$$\begin{aligned} e^{a+b} &= e^a e^b \textit{ for all } a, b \in \mathbf{R}. \\ e^{a-b} &= \frac{e^a}{e^b} \textit{ for all } a, b \in \mathbf{R}. &\quad (14.32) \\ (e^a)^q &= e^{aq} \textit{ for all } a \in \mathbf{R}, \textit{ and for all } q \in \mathbf{Q}. \\ (e^a)^{-1} &= e^{-a} \textit{ for all } a \in \mathbf{R}. \\ e^{\ln(x)} &= x \textit{ for all } x \in \mathbf{R}^+. \\ \ln(e^a) &= a \textit{ for all } a \in \mathbf{R}. \\ e^0 &= 1. \\ e^1 &= e. &\quad (14.33) \end{aligned}$$

Proof: We have proved all of these properties except for relation (14.32). The proof of (14.32) is the next exercise.

14.34 Exercise. Show that $e^{a-b} = \dfrac{e^a}{e^b}$ for all $a, b \in \mathbf{R}$.

14.35 Exercise. Show that if $a \in \mathbf{R}^+$ and $q \in \mathbf{Q}$, then

$$a^q = e^{q\ln(a)}.$$

14.36 Definition (a^x.) The result of the last exercise motivates us to make the definition

$$a^x = e^{x\ln(a)} \text{ for all } x \in \mathbf{R} \text{ and for all } a \in \mathbf{R}^+.$$

14.37 Exercise. Prove the following results:

$$\begin{aligned} a^x a^y &= a^{x+y} \text{ for all } a \in \mathbf{R}^+ \text{ and for all } x, y \in \mathbf{R}. \\ (a^x)^y &= a^{xy} \text{ for all } a \in \mathbf{R}^+ \text{ and for all } x, y \in \mathbf{R}. \\ (ab)^x &= a^x b^x \text{ for all } a, b \in \mathbf{R}^+ \text{ and for all } x \in \mathbf{R}. \\ \ln(a^x) &= x\ln(a) \text{ for all } a \in \mathbf{R}^+ \text{ and for all } x \in \mathbf{R}. \end{aligned}$$

14.5 Inverse Function Theorems

14.38 Lemma. *Let f be a strictly increasing continuous function whose domain is an interval $[a, b]$. Then the image of f is the interval $[f(a), f(b)]$, and the function $f : [a, b] \to [f(a), f(b)]$ has an inverse.*

Proof: It is clear that $f(a)$ and $f(b)$ are in image(f). Since f is continuous we can apply the intermediate value property to conclude that for every number z between $f(a)$ and $f(b)$ there is a number $c \in [a, b]$ such that $z = f(c)$, i.e. $[f(a), f(b)] \subset \text{image}(f)$. Since f is increasing on $[a, b]$ we have $f(a) \le f(t) \le f(b)$ whenever $a \le t \le b$, and thus $\text{image}(f) \subset [f(a), f(b)]$. It follows that $f : [a, b] \to [f(a), f(b)]$ is surjective, and since strictly increasing functions are injective, f is bijective. By remark (14.22) f has an inverse.

14.39 Exercise. State and prove the analogue of lemma 14.38 for strictly decreasing functions.

14.40 Exercise. Let f be a function whose domain is an interval $[a, b]$, and whose image is an interval. Does it follow that f is continuous?

14.41 Exercise. Let f be a continuous function on a closed bounded interval $[a, b]$. Show that the image of f is a closed bounded interval $[A, B]$.

14.42 Exercise. Let J and I be non-empty intervals and let $f : J \to I$ be a continuous function such that $I = \text{image}(f)$.

a) Show that if f is strictly increasing, then the inverse function for f is also strictly increasing.

b) Show that if f is strictly decreasing, then the inverse function for f is also strictly decreasing.

14.43 Theorem (Inverse function theorem.) *Let f be a continuous strictly increasing function on an interval $J = [a, b]$ of positive length, such that $f'(x) > 0$ for all $x \in \text{interior}(J)$. Let I be the image of J and let*

$$g : I \to J$$

be the inverse function for f. Then g is differentiable on the interior of I and

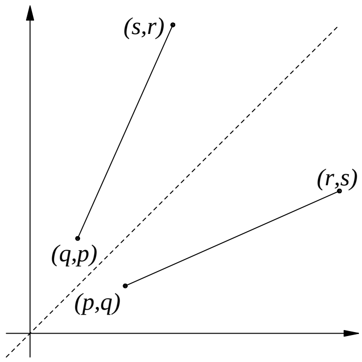
$$g'(s) = \frac{1}{f'(g(s))} \text{ for all } s \in \text{interior}(I) \tag{14.44}$$

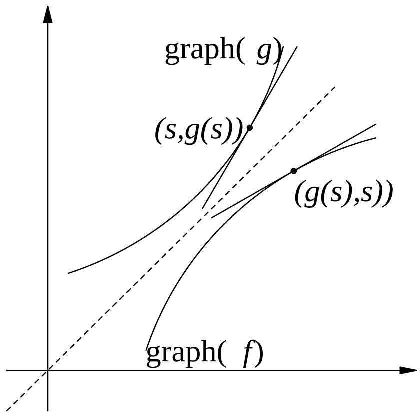
Remark: If l is a nonvertical line joining two points (p, q) and (r, s) then the slope of l is

$$m = \frac{s - q}{r - p}.$$

The reflection of l about the line whose equation is $y = x$ passes through the points (q, p) and (s, r), so the slope of the reflected line is

$$\frac{r - p}{s - q} = \frac{1}{m}.$$





$$g'(s) = \frac{1}{f'(g(s))}$$

Thus theorem 14.43 says that the tangent to graph(g) at the point $(s, g(s))$ is obtained by reflecting the tangent to graph(f) at $(g(s), s)$ about the line whose equation is $y = x$. This is what you should expect from the geometry of the situation.

Proof of theorem 14.43: The first thing that should be done, is to prove that g is continuous. I am going to omit that proof and just assume the continuity of g, and then show that g is differentiable, and that g' is given by formula (14.44).

Let s be a point in the interior of dom(g). then

$$\begin{aligned} \lim_{t \to s} \frac{g(t) - g(s)}{t - s} &= \lim_{t \to s} \frac{g(t) - g(s)}{f(g(t)) - f(g(s))} \\ &= \lim_{t \to s} \frac{1}{\frac{f(g(t)) - f(g(s))}{g(t) - g(s)}}. \end{aligned} \tag{14.45}$$

(Observe that we have not divided by zero). Let $\{t_n\}$ be a sequence in $\text{dom}(g) \setminus \{s\}$ such that $\{t_n\} \to s$. Then $\{g(t_n)\} \to g(s)$ (since g is assumed to be continuous), and $g(t_n) \neq g(s)$ for all $n \in \mathbf{Z}^+$ (since g is injective). Since f is differentiable at $g(s)$, it follows that

$$\left\{ \frac{f(g(t_n)) - f(g(s))}{g(t_n) - g(s)} \right\} \to f'(g(s)).$$

Since $f'(g(s)) \neq 0$ it follows that

$$\left\{ \frac{1}{\frac{f(g(t_n))-f(g(s))}{g(t_n)-g(s)}} \right\} \to \frac{1}{f'(g(s)}.$$

It follows that

$$\lim_{t \to s} \frac{g(t) - g(s)}{t - s} = \frac{1}{f'(g(s))}$$

and the theorem is proved. ‖

Remark: The inverse function theorem also applies to continuous functions f on J such that $f'(s) < 0$ for all $s \in$ interior (a, b). Formula (14.44) is valid in this case also.

Remark: Although we have stated the inverse function theorem for functions on intervals of the form $[a, b]$, it holds for functions defined on any interval. Let J be an interval, and let f be a continuous strictly increasing function from J to $\mathbf{R}$ such that $f'(x) > 0$ for all x in the interior of J. Let p be a point in the interior of image(J). Then we can find points r and s in image(J) such that $r < p < s$. Now f maps the interval $[g(r), g(s)]$ bijectively onto $[r, s]$, and since $p \in (r, s)$ we can apply the inverse function theorem on the interval $[g(r), g(s)]$ to conclude that $g'(p) = \dfrac{1}{f'(g(p))}$. It is not necessary to remember the formula for $g'(p)$. Once we know that g is differentiable, we can calculate g' by using the chain rule, as illustrated by the examples in the next section.

14.6 Some Derivative Calculations

14.46 Example (Derivative of exp.**)** We know that

$$\ln(E(t)) = t \text{ for all } t \in \mathbf{R}.$$

If we differentiate both sides of this equation, we get

$$\frac{1}{E(t)}E'(t) = 1,$$

i.e.

$$E'(t) = E(t) \text{ for all } t \in \mathbf{R}.$$

14.47 Example (Derivative of x^r.) Let r be any real number and let $f(x) = x^r$ for all $x \in \mathbf{R}^+$. Then

$$f(x) = x^r = E(r\ln(x)),$$

so by the chain rule

$$f'(x) = E'(r\ln(x)) \cdot \frac{r}{x} = E(r\ln(x)) \cdot rx^{-1} = x^r rx^{-1} = rx^{r-1}.$$

(Here I have used the result of exercise 14.37.) Thus the formula

$$\frac{d}{dx}(x^r) = rx^{r-1}$$

which we have known for quite a while for rational exponents, is actually valid for all real exponents.

14.48 Exercise (Derivative of a^x.) Let $a \in \mathbf{R}^+$. Show that

$$\frac{d}{dx}(a^x) = a^x \ln(a)$$

for all $x \in \mathbf{R}$.

14.49 Example (Derivative of x^x.)

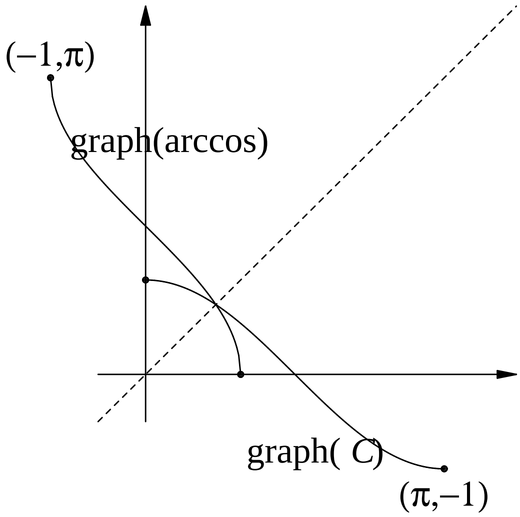
$$\begin{aligned}\frac{d}{dx}x^x &= \frac{d}{dx}e^{x\ln(x)} = e^{x\ln(x)}\frac{d}{dx}(x\ln(x)) \\ &= x^x(x \cdot \frac{1}{x} + \ln(x)) = x^x(1 + \ln(x)).\end{aligned}$$

Hence

$$\frac{d}{dx}x^x = x^x(1 + \ln(x)) \text{ for all } x \in \mathbf{R}^+.$$

14.50 Example (Derivative of arccos**.)** Let $C : [0, \pi] \to [-1, 1]$ be defined by

$$C(x) = \cos(x) \text{ for all } x \in [0, \pi].$$



We have

$$C'(x) = -\sin(x) < 0 \text{ for all } x \in (0, \pi),$$

so C has an inverse function which is denoted by arccos. By the inverse function theorem arccos is differentiable on $(-1, 1)$. and we have

$$\cos(\arccos(t)) = C(\arccos(t)) = t \text{ for all } t \in [-1, 1].$$

By the chain rule

$$-\sin(\arccos(t))\arccos'(t) = 1 \text{ for all } t \in (-1, 1).$$

Now since the sine function is positive on $(0, \pi)$ we get

$$\sin(s) = \sqrt{1 - \cos^2(s)}$$

for all $s \in (0, \pi)$, so

$$\sin(\arccos(t)) = \sqrt{1 - (\cos(\arccos(t)))^2} = \sqrt{1 - t^2} \text{ for all } t \in (-1, 1).$$

Thus

$$\arccos'(t) = \frac{-1}{\sin(\arccos(t))} = \frac{-1}{\sqrt{1 - t^2}} \text{ for all } t \in (-1, 1).$$

14.51 Exercise (Derivative of arcsin.**)** Let

$$S(t) = \sin(t) \text{ for all } t \in [-\frac{\pi}{2}, \frac{\pi}{2}].$$

Show that S has an inverse function that is differentiable on the interior of its domain. This inverse functions is called arcsin. Describe the domain of arcsin, sketch the graphs of S and of arcsin, and show that

$$\frac{d}{dx}\arcsin(x) = \frac{1}{\sqrt{1-x^2}}.$$

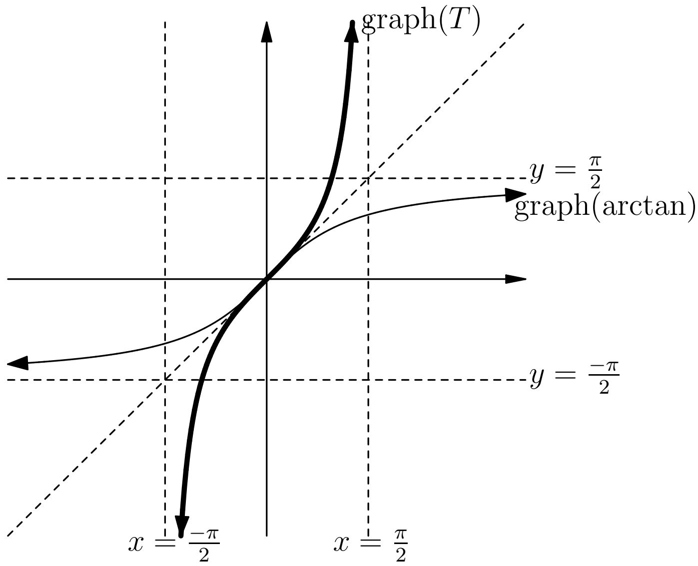
14.52 Example (Derivative of arctan.**)** Let

$$T(x) = \tan(x) \text{ for all } x \in (-\frac{\pi}{2}, \frac{\pi}{2}).$$

Then T is continuous, and the image of T is unbounded both above and below, so image$(T) = \mathbf{R}$. Also

$$T'(x) = \sec^2(x) > 0 \text{ for all } x \in (-\frac{\pi}{2}, \frac{\pi}{2})$$

so T has an inverse function, which we denote by arctan.



For all $x \in \mathbf{R}$

$$\tan(\arctan(x)) = T(\arctan(x)) = x,$$

so by the chain rule

$$\sec^2(\arctan(x))\arctan'(x) = 1 \text{ for all } x \in \mathbf{R}.$$

Now

$$\sec^2(t) = 1 + \tan^2(t) \text{ for all } t \in \mathrm{dom}(\sec),$$

so

$$\sec^2(\arctan(x)) = 1 + \tan^2(\arctan(x)) = 1 + x^2 \text{ for all } x \in \mathbf{R}.$$

Thus

$$\arctan'(x) = \frac{1}{\sec^2(\arctan(x))} = \frac{1}{1+x^2} \text{ for all } x \in \mathbf{R}.$$

14.53 Exercise (Derivative of arccot**.)** Let

$$V(x) = \cot(x) \text{ for all } x \in (0, \pi).$$

Show that V has an inverse function arccot, and that

$$\frac{d}{dx}\mathrm{arccot}(x) = -\frac{1}{1+x^2}.$$

What is dom(arccot)? Sketch the graphs of V and of arccot.

Remark The first person to give a name to the inverse trigonometric functions was Daniel Bernoulli (1700-1792) who used AS for arcsin in 1729. Other early notations included arc(cos. $= x$) and ang(cos. $= x$)[15, page 175]. Many calculators and some calculus books use $\cos^{-1}$ to denote arccos. (If you use your calculator to find inverse trigonometric functions, make sure that you set the degree-radian-grad mode to radians.)

14.54 Exercise. Calculate the derivatives of the following functions, and simplify your answers (Here a is a constant.)

a) $f(x) = x\sqrt{a^2 - x^2} + a^2 \arcsin(\frac{x}{a})$.

b) $g(x) = \arcsin(x) + \frac{x}{1-x^2}$.

c) $h(x) = x \arccos(ax) - \dfrac{1}{a}\sqrt{1 - a^2x^2}$.

d) $k(x) = \arctan(e^x + e^{-x})$.

e) $m(x) = x\sqrt{1 - x^2} + \arcsin(x)(2x^2 - 1)$.

f) $n(x) = e^{ax}(a\sin(bx) - b\cos(bx))$. Here a and b are constants.

g) $p(x) = e^{ax}(a^2x^2 - 2ax + 2)$. Here a is a constant..

14.55 Exercise. Let

$$l(x) = \arctan(\tan(x)).$$

Calculate the derivative of l. What is the domain of this function? Sketch the graph of l.

14.56 Exercise (Hyperbolic functions.) We define functions sinh and cosh on $\mathbf{R}$ by

$$\begin{aligned} \cosh(x) &= \frac{e^x + e^{-x}}{2} \text{ for all } x \in \mathbf{R}. \\ \sinh(x) &= \frac{e^x - e^{-x}}{2} \text{ for all } x \in \mathbf{R}. \end{aligned}$$

These functions are called the *hyperbolic sine* and the *hyperbolic cosine* respectively. Show that

$$\frac{d}{dx}\cosh(x) = \sinh(x),$$

and

$$\frac{d}{dx}\sinh(x) = \cosh(x).$$

Calculate

$$\frac{d}{dx}\Big(\cosh^2(x) - \sinh^2(x)\Big),$$

and simplify your answer as much as you can. What conclusion can you draw from your answer? Sketch the graphs of cosh and sinh on one set of coordinate axes.

Chapter 15

The Second Derivative

15.1 Higher Order Derivatives

15.1 Definition (Higher order derivatives.) Let f be a function whose domain is a subset of $\mathbf{R}$. We define a function f' (called the *derivative of f*) by

$$\text{domain}(f') = \{x \in \text{dom}(f) : f'(x) \text{ exists}\}.$$

and for all $x \in \text{dom}(f)$, the value of f' at x is the derivative $f'(x)$. We may also write $f^{(1)}$ for f'. Since f' is itself a function, we can calculate its derivative: this derivative is denoted by f'' or $f^{(2)}$, and is called the *second derivative of f*. For integers $n \geq 2$ we define

$$f^{(n+1)} = (f^{(n)})'. \tag{15.2}$$

and we call $f^{(n)}$ the *nth derivative of f*. We also define

$$f^{(0)} = f.$$

In Leibniz's notation we write

$$\frac{d^n f}{dx^n} = f^{(n)}, \text{ or } \frac{d^n}{dx^n} f = f^{(n)}, \text{ or } \left(\frac{d}{dx}\right)^{(n)} f(x) = f^{(n)}(x) \text{ or } \frac{d^n f}{dx^n} = f^{(n)}(x),$$

so that equation (15.2) becomes

$$\frac{d^{n+1} f}{dx^{n+1}} = \frac{d}{dx}\left(\frac{d^n f}{dx^n}\right).$$

If a and b are real numbers, and f and g are functions then from known properties of the derivative we can show that

$$(af+bg)^{(n)} = af^{(n)} + bg^{(n)} \text{ on } \mathrm{dom}(f^{(n)}) \cap \mathrm{dom}(g^{(n)}).$$

or

$$\frac{d^n}{dx^n}(af+bg) = a\frac{d^n f}{dx^n} + b\frac{d^n f}{dx^n}.$$

15.3 Examples. If $h(x) = \sin(\omega x)$, where $\omega \in \mathbf{R}$, then

$$\begin{aligned} h'(x) &= \omega\cos(\omega x), \\ h''(x) &= -\omega^2\sin(\omega x), \\ h^{(3)}(x) &= -\omega^3\cos(\omega x), \\ h^{(4)}(x) &= \omega^4\sin(\omega x) = \omega^4 h(x). \end{aligned}$$

It should now be apparent that

$$h^{(4n+k)}(x) = \omega^{4n}h^{(k)}(x) \text{ for } k = 0,1,2,3.$$

so that

$$h^{(98)}(x) = h^{(4\cdot 24+2)}(x) = \omega^{96}h^{(2)}(x) = -\omega^{98}\sin(\omega x).$$

If

$$g(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

then

$$\begin{aligned} g'(x) &= 1 + x + \frac{x^2}{2!}, \\ g''(x) &= 1 + x, \\ g^{(3)}(x) &= 1, \\ g^{(n)}(x) &= 0 \text{ for } n \in \mathbf{Z}_{\geq 4}. \end{aligned}$$

If $y = \ln(x)$ then

$$\begin{aligned} \frac{dy}{dx} &= \frac{1}{x}, \\ \frac{d^2y}{dx^2} &= -\frac{1}{x^2}, \\ \frac{d^3y}{dx^3} &= \frac{2}{x^3}. \end{aligned}$$

15.4 Exercise. Calculate $g^{(5)}(t)$ if $g(t) = t^4 \ln(t)$.

15.5 Exercise. Let $g(t) = tf(t)$. Calculate $g'(t)$, $g''(t)$, $g^{(3)}(t)$ and $g^{(4)}(t)$ in terms of $f(t)$, $f'(t)$, $f''(t)$, $f^{(3)}(t)$ and $f^{(4)}(t)$. What do you think is the formula for $g^{(n)}(t)$?

15.6 Exercise. Find $\dfrac{d^2y}{dx^2}$ if $y = 1/(x^2 - 1)$.

15.7 Exercise. Find $f''(x)$ if $f(x) = e^{\frac{1}{x^2}} = \exp\left(\dfrac{1}{x^2}\right)$.

15.8 Exercise. Suppose $f''(x) = 0$ for all $x \in \mathbf{R}$. What can you say about f?

15.9 Exercise. Let f and g be functions such that $f^{(2)}$ and $g^{(2)}$ are defined on all of $\mathbf{R}$. Show that

$$(fg)^{(2)} = fg^{(2)} + 2f^{(1)}g^{(1)} + f^{(2)}g.$$

Find a similar function for $(fg)^{(3)}$ (assuming that $f^{(3)}$ and $g^{(3)}$ are defined.)

In Leibniz's calculus, d^2f or ddf was actually an infinitely small quantity that was so much smaller than dx that the quotient $\dfrac{d^2f}{dx}$ was zero, and $\dfrac{d^2f}{dx^2}$ was obtained by multiplying dx by itself and then dividing the result into d^2f. Leibniz also used notations like $\dfrac{ddy}{ddx}$ and $\dfrac{dxds}{ddy}$ for which our modern notation has no counterparts. Leibniz considered the problem of defining a meaning for $d^{\frac{1}{2}}f$, but he did not make much progress on this problem. Today there is considerable literature on fractional derivatives. A brief history of the subject can be found in [36, ch I and ch VIII].

15.10 Exercise. Let a be a real number. Show that for $k = 0, 1, 2, 3$

$$\frac{d^k}{dx^k} e^{ax} = a^k e^{ax}. \tag{15.11}$$

After doing this it should be clear that equation (15.11), in fact holds for all $n \in \mathbf{Z}_{\geq 0}$ (this can be proved by induction). Now suppose that $a > 0$ and we will *define*

$$\frac{d^r}{dx^r} e^{ax} = a^r e^{ax} \text{ for all } r \in \mathbf{R}. \tag{15.12}$$

Show that then for all p and q in $\mathbf{R}$,

$$\left(\frac{d}{dx}\right)^p \left(\left(\frac{d}{dx}\right)^q (e^{ax})\right) = \left(\frac{d}{dx}\right)^{p+q} (e^{ax}).$$

Find $\left(\frac{d}{dx}\right)^{\frac{1}{2}} e^{3x}$ and $\left(\frac{d}{dx}\right)^{\frac{1}{2}} e^{5x}$. What do you think $\left(\frac{d}{dx}\right)^{\frac{1}{2}} \left(3e^{3x} + 4e^{5x}\right)$ should be?

Equation (15.12) was the starting point from which Joseph Liouville (1809–1882) developed a theory of fractional calculus[36, pp 4-6].

15.13 Exercise. Let a and b be real numbers. Show that for $k = 0, 1, 2, 3$

$$\left(\frac{d}{dx}\right)^k \cos(ax + b) = a^k \cos(ax + b + \frac{k\pi}{2}). \tag{15.14}$$

After doing this exercise it should be clear that in fact equation (15.14) holds for all $k \in \mathbf{Z}_{\geq 0}$ (this can be proved by induction). Now suppose that $a > 0$, and we will *define*

$$\left(\frac{d}{dx}\right)^r \cos(ax + b) = a^r \cos(ax + b + \frac{r\pi}{2}) \text{ for all } r \in \mathbf{R}. \tag{15.15}$$

Show that for all p and q in $\mathbf{R}$

$$\left(\frac{d}{dx}\right)^p \left(\left(\frac{d}{dx}\right)^q \cos(ax + b)\right) = \left(\frac{d}{dx}\right)^{p+q} \cos(ax + b).$$

Equation (15.15) was used as the starting point for a definition of fractional derivatives for general functions, by Joseph Fourier (1768–1830)[36, page 3].

15.2 Acceleration

15.16 Definition (Acceleration.) If a particle **p** moves in a straight line so that its position at time t is $h(t)$, we have defined its velocity at time t to be $h'(t)$. We now define its *acceleration* at time t to be $h''(t)$, so that acceleration is the derivative of velocity. Thus if a particle moves with a constant acceleration of $1\frac{\text{ft./sec.}}{\text{sec.}}$, then every second its velocity increases by one ft./sec.

15.17 Example. A mass on the end of a spring moves so that its height at time t is $-A\cos(\omega t)$, where A and ω are positive numbers. If we denote its velocity at time t by $v(t)$, and its acceleration at time t by $a(t)$ then

$$\begin{aligned} h(t) &= -A\cos(\omega t) \\ v(t) = h'(t) &= A\omega\sin(\omega t) \\ a(t) = v'(t) &= A\omega^2\cos(\omega t) \end{aligned}$$

From this we see that the acceleration is always of opposite sign from the position: when the mass is above the zero position it is being accelerated downward, and when it is below its equilibrium position it is being accelerated upward. Also we see that the magnitude of the acceleration is largest when the velocity is 0.

15.18 Definition (Acceleration due to gravity.) If a particle **p** moves near the surface of the earth, acted on by no forces except the force due to gravity, then **p** will move with a constant acceleration $-g$ which is independent of the mass of **p**. The value of g is

$$g = \frac{32\text{ft./sec.}}{\text{sec.}}(\text{approx.}) \quad \text{or} \quad g = \frac{9.8\text{meter/sec.}}{\text{sec.}}(\text{approx.}).$$

We call g *the acceleration due to gravity.* Actually, the value of g varies slightly over the surface of the earth, so there is no exact value for g. The law just described applies in situations when air resistance and buoyancy can be neglected. It describes the motion of a falling rock well, but it does not describe a falling balloon.

Remark: When I solve applied problems, I will usually omit all units (e.g. feet or seconds) in my calculations, and will put them in only in the final answers.

15.19 Example. A juggler J tosses a ball vertically upward from a height of 4 feet above the ground with a speed of 16 ft./sec. Let $h(t)$ denote the height of the ball above the ground at time t. We will set our clock so that $t = 0$ corresponds to the time of the toss:

$$h(0) = 4; \qquad h'(0) = 16.$$

We will suppose that while the ball is in the air, its motion is described by a differentiable function of t. We assume that

$$h''(t) = -g = -32.$$

We know one function whose derivative is $-g$:

$$\text{if } s(t) = -gt, \text{ then } s'(t) = -g.$$

By the antiderivative theorem it follows that there is a constant v_0 such that

$$h'(t) = s(t) + v_0 = -gt + v_0.$$

Moreover we can calculate v_0 as follows:

$$(16 = h'(0) = -g \cdot 0 + v_0) \Longrightarrow (v_0 = 16).$$

Thus

$$h'(t) = -gt + 16.$$

We know a function whose derivative is $-gt + 16$:

$$\text{if } w(t) = -\frac{gt^2}{2} + 16t, \text{ then } w'(t) = -gt + 16.$$

Thus there is a constant h_0 such that

$$h(t) = w(t) + h_0 = -\frac{gt^2}{2} + 16t + h_0.$$

To find h_0 we set $t = 0$:

$$(4 = h(0) = -\frac{g \cdot 0^2}{2} + 16 \cdot 0 + h_0) \Longrightarrow (h_0 = 4).$$

Thus

$$h(t) = -\frac{gt^2}{2} + 16t + 4.$$

The ball will reach its maximum height when $h'(t) = 0$, i.e. when

$$t = \frac{16}{g} = \frac{16}{32} = \frac{1}{2}.$$

The maximum height reached by the ball is

$$h(\frac{1}{2}) = -\frac{1}{2} \cdot 32 \cdot (\frac{1}{2})^2 + \frac{16}{2} + 4 = 8,$$

so the ball rises to a maximum height of 8 feet above the ground.

15.20 Example (Conservation of energy.) Suppose that a particle $\mathbf{p}$ moves near the surface of the earth acted upon by no forces except the force of gravity. Let $v(t)$ and $h(t)$ denote respectively its height above the earth and its velocity at time t. Then

$$\frac{dv}{dt} = h''(t) = -g,$$

so

$$v\frac{dv}{dt} = -gv = -g\frac{dh}{dt}.$$

Now

$$v\frac{dv}{dt} = \frac{d}{dt}(\frac{1}{2}v^2),$$

so we have

$$\frac{d}{dt}(\frac{1}{2}v^2) = \frac{d}{dt}(-gh).$$

It follows that there is a constant K such that

$$\frac{1}{2}v^2 = -gh + K,$$

or

$$\frac{1}{2}v^2 + gh = K.$$

If m is the mass of the particle $\mathbf{p}$ then

$$\frac{1}{2}mv^2 + mgh = Km. \tag{15.21}$$

The quantity $\frac{1}{2}mv^2$ is called the *kinetic energy of* $\mathbf{p}$, and the quantity mgh is called the *potential energy* of $\mathbf{p}$. Equation (15.21) states that as $\mathbf{p}$ moves, the sum of its potential energy end its kinetic energy remains constant.

15.22 Exercise. A particle moves in a vertical line near the surface of the earth, acted upon by no forces except the force of gravity. At time 0 it is at height h_0, and has velocity v_0. Derive the formula for the height of the particle at time $t > 0$.

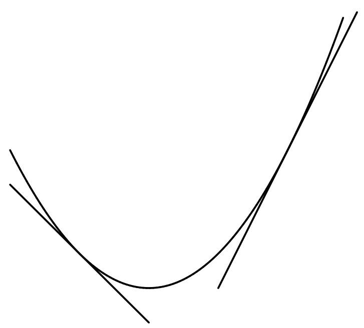
15.23 Exercise. The acceleration due to gravity on the moon is approximately

$$g_m = .17g$$

where g denotes the acceleration due to gravity on the earth. A juggler J on the moon wants to toss a ball vertically upward so that it rises 4 feet above its starting height. With what velocity should the ball leave J's hand?

15.3 Convexity

15.24 Definition (Convexity) Let f be a differentiable function on an interval (a, b). We say that f is *convex upward over* (a, b) or that f *holds water over* (a, b) if and only if for each point t in (a, b), the tangent line to graph(f) at $(t, f(t))$ lies below the graph of f.



convex upward curve (holds water)

Since the equation of the tangent line to graph(f) at $(t, f(t))$ is

$$y = f(t) + f'(t)(x - t),$$

the condition for f to be convex upward over (a, b) is that for all x and t in (a, b)

$$f(t) + f'(t)(x - t) \leq f(x). \tag{15.25}$$

Condition (15.25) is equivalent to the two conditions:

$$f'(t) \leq \frac{f(x) - f(t)}{x - t} \text{ whenever } t < x,$$

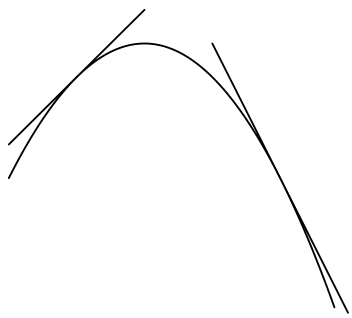
and

$$\frac{f(t) - f(x)}{t - x} \leq f'(t) \text{ whenever } x < t.$$

These last two conditions can be written as the single condition

$$f'(p) \leq \frac{f(q) - f(p)}{q - p} \leq f'(q) \text{ whenever } p < q. \tag{15.26}$$

We say that f is *convex downward over* (a, b), or that f *spills water over* (a, b) if and only if for each point t in (a, b), the tangent line to graph(f) at $(t, f(t))$ lies above the graph of f.



convex downward curve (spills water)

This condition is equivalent to the condition that for all points $p, q \in (a, b)$

$$f'(p) \geq \frac{f(q) - f(p)}{q - p} \geq f'(q) \text{ whenever } p < q.$$

15.27 Theorem. *Let f be a differentiable function over the interval (a, b). Then f is convex upward over (a, b) if and only if f' is increasing over (a, b). (and similarly f is convex downward over (a, b) if and only if f' is decreasing over (a, b).)*

Proof: If f is convex upward over (a, b), then it follows from (15.26) that f' is increasing over (a, b).

Now suppose that f' is increasing over (a,b). Let p,q be distinct points in (a,b). By the mean value theorem there is a point c between p and q such that

$$f'(c) = \frac{f(p)-f(q)}{p-q}.$$

If $p < q$ then $p < c < q$ so since f' is increasing over (a,b)

$$f'(p) \leq f'(c) \leq f'(q),$$

i.e.

$$f'(p) \leq \frac{f(p)-f(q)}{p-q} \leq f'(q).$$

Thus condition (15.26) is satisfied, and f is convex upward over (a,b).

15.28 Corollary. *Let f be a function such that $f''(x)$ exists for all x in the interval (a,b). If $f''(x) \geq 0$ for all $x \in (a,b)$ then f is convex upward over (a,b). If $f''(x) \leq 0$ for all $x \in (a,b)$ then f is convex downward over (a,b).*

15.29 Exercise. Prove one of the two statements in corollary 15.28.

15.30 Lemma (Converse of corollary 12.26) *Let f be a real function such that f is continuous on $[a,b]$ and differentiable on (a,b). If f is increasing on $[a,b]$, then $f'(x) \geq 0$ for all $x \in (a,b)$.*

Proof: let $p \in (a,b)$. Choose $\delta > 0$ such that $(p-\delta, p+\delta) \subset (a,b)$. Then $\{p + \frac{\delta}{2n}\}$ is a sequence such that

$$\{p + \frac{\delta}{2n}\} \to p,$$

and hence

$$\{\frac{f(p+\frac{\delta}{2n}) - f(p)}{(p+\frac{\delta}{2n}) - p}\} \to f'(p).$$

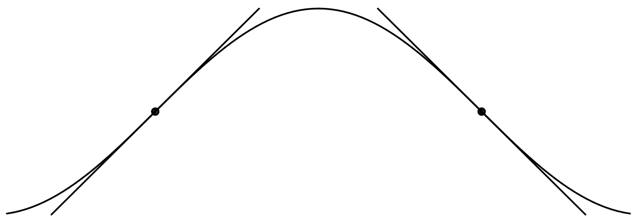
Since f is increasing on (a,b), we have

$$\frac{f(p+\frac{\delta}{2n}) - f(p)}{(p+\frac{\delta}{2n}) - p} \geq 0$$

for all $n \in \mathbf{Z}^+$, and it follows that

$$f'(p) \geq 0 \text{ for all } p \in (a,b). \; \|$$

15.31 Definition (Inflection point) Let f be a real function, and let $a \in \text{dom} f$. We say that a is a *point of inflection* for f if there is some $\epsilon > 0$ such that $(a-\epsilon, a+\epsilon) \subset \text{dom} f$, and f is convex upward on one of the intervals $(a-\epsilon, a)$, $(a, a+\epsilon)$, and is convex downward on the other.



15.32 Theorem (Second derivative test for inflection points) *Let f be a real function, and let a be a point of inflection for f. If f'' is defined and continuous in some interval $(a-\delta, a+\delta)$ then $f''(a) = 0$.*

Proof: We will suppose that f is convex upward on the interval $(a-\delta, a)$ and is convex downward on $(a, a+\delta)$. (The proof in the case where these conditions are reversed is essentially the same). Then f' is increasing on $(a-\delta, a)$, and f' is decreasing on $(a, a+\delta)$. By (15.30), $f''(x) \geq 0$ for all $x \in (a-\delta, a)$, and $f''(x) \leq 0$ for all $x \in (a, a+\delta)$. We have

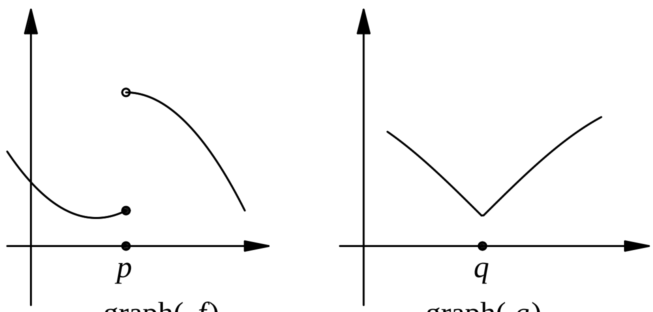
$$f''(a) = \lim\{f''(a + \frac{\delta}{2n})\} \leq 0,$$

and

$$f''(a) = \lim\{f''(a - \frac{\delta}{2n})\} \geq 0.$$

It follows that $f''(a) = 0$. |||

15.33 Example. When you look at the graph of a function, you can usually "see" the points where the second derivative changes sign. However, most people cannot "see" points where the second derivative is undefined.



By inspecting graph(f), you can see that f has a discontinuity at p.

By inspecting graph(g), you can see that g is continuous everywhere, but g' is not defined at q.

By inspecting graph(h) in figure a below, you can see that h' is continuous, but you may have a hard time seeing the point where h'' is not defined.

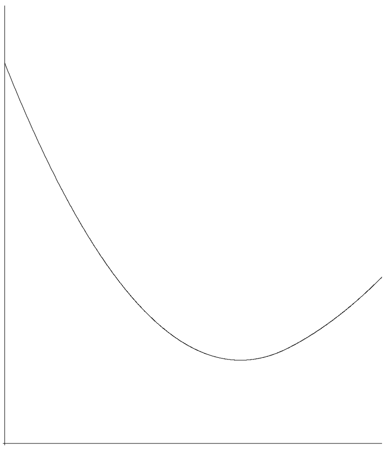


figure a graph(h)

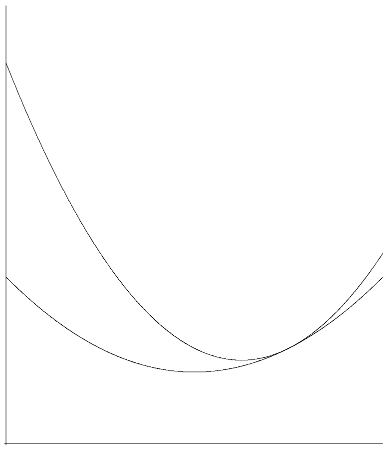


figure b

The function h is defined by

$$h(x) = \begin{cases} x^2 - \frac{5}{2}x + 2 & \text{if } 0 \leq x \leq \frac{3}{2}. \\ \frac{1}{2}x^2 - x + \frac{7}{8} & \text{if } \frac{3}{2} < x \leq 2. \end{cases} \tag{15.34}$$

so $h''(x) = 2$ for $0 < x < \frac{3}{2}$, and $h''(x) = 1$ for $\frac{3}{2} < x < 2$, and $h''(\frac{3}{2})$ is not defined. We constructed h by pasting together two parabolas. Figure b shows the two parabolas, one having a second derivative equal to 1, and the other having second derivative equal to 2.

15.35 Exercise. Let h be the function described in formula (15.34). Draw graphs of h' and h''.

15.36 Entertainment (Discontinuous derivative problem.) There exists a function f such that f is differentiable everywhere on $\mathbf{R}$, but f' is discontinuous somewhere. Find such a function.

15.37 Exercise. Let $f(x) = x^4$. Show that $f''(0) = 0$, but 0 is not a point of inflection for f. Explain why this result does not contradict theorem 15.32

15.38 Example. Let

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

Then

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

and

$$f''(x) = \frac{(1+x^2)^2(-2) - (-2x)(2(1+x^2)(2x))}{(1+x^2)^4} = \frac{2(3x^2-1)}{(1+x^2)^3}.$$

Thus the only critical point for f is 0. Also,

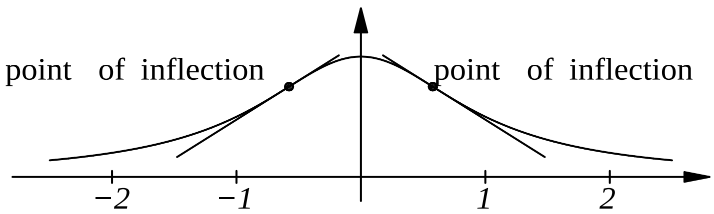
$$(f'(x) > 0 \iff x < 0) \text{ and } (f'(x) < 0 \iff x > 0),$$

so f is increasing on $(-\infty, 0)$ and is decreasing on $(0, \infty)$. Thus f has a maximum at 0, and f has no minima.

We see that $f''(x) = 0 \iff x^2 = \frac{1}{3}$, and moreover

$$(f''(x) < 0) \iff x \in \left(-\sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}}\right),$$

so f spills water over the interval $\left(-\sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}}\right)$, and f holds water over each of the intervals $\left(-\infty, -\sqrt{\frac{1}{3}}\right)$ and $\left(\sqrt{\frac{1}{3}}, \infty\right)$. Thus f has points of inflection at $\pm\sqrt{\frac{1}{3}}$. We can use all of this information to make a reasonable sketch of the graph of f. Note that $f(x) > 0$ for all x, $f(0) = 1$, and $f\left(\pm\sqrt{\frac{1}{3}}\right) = \frac{3}{4}$, and $\sqrt{\frac{1}{3}}$ is approximately 0.58.



$$\left\{ y = \frac{1}{1+x^2} \right\}$$

15.39 Exercise. Discuss the graphs of the following functions. Make use of all the information that you can get by looking at the functions and their first two derivatives.

a) $f(x) = 5x^4 - 4x^5$.
b) $G(x) = 5x^3 - 3x^5$.
c) $H(x) = e^{-\frac{1}{x^2}}$.

Chapter 16

Fundamental Theorem of Calculus

16.1 Definition (Nice functions.) I will say that a real valued function f defined on an interval $[a, b]$ is a *nice function* on $[a, b]$, if f is continuous on $[a, b]$ and integrable on every subinterval of $[a, b]$.

Remark: We know that piecewise monotonic continuous functions on $[a, b]$ are nice. It turns out that every continuous function on $[a, b]$ is nice, but we are not going to prove this. The theorems stated in this chapter for nice functions are usually stated for continuous functions. You can find a proof that every continuous function on an interval $[a, b]$ is integrable on $[a, b]$ (and hence that every continuous function on $[a, b]$ is nice on $[a, b]$) in [44, page 246] or in [1, page 153]. However both of these sources use a slightly different definition of continuity and of integral than we do, so you will need to do some work to translate the proofs in these references into proofs in our terms. You might try to prove the result yourself, but the proof is rather tricky. For all the applications we will make in this course, the functions examined will be continuous and piecewise monotonic so the theorems as we prove them are good enough.

16.2 Exercise. Can you give an example of a continuous function on a closed interval that is *not* piecewise monotonic? You may describe your example rather loosely, and you do not need to prove that it is continuous.

16.3 Theorem (Fundamental theorem of calculus I.) *Let g be a nice function on $[a, b]$. Suppose G is an antiderivative for g on $[a, b]$. Then G is* *an indefinite integral for g on $[a,b]$; i.e.,*

$$\int_p^q g = G(q) - G(p) \text{ for all } p, q \in [a,b]. \tag{16.4}$$

Proof: By the definition of antiderivative, G is continuous on $[a,b]$ and $G' = g$ on (a,b). Let p,q be arbitrary points in $[a,b]$. I will suppose $p < q$. (Note that if (16.4) holds when $p < q$, then it holds when $q < p$, since both sides of the equation change sign when p and q are interchanged. Also note that the theorem clearly holds for $p = q$.)

Let $P = \{x_0, \cdots, x_m\}$ be any partition of $[p,q]$, and let i be an integer with $1 \leq i \leq m$. If $x_{i-1} < x_i$ we can apply the mean value theorem to G on $[x_{i-1}, x_i]$ to find a number $s_i \in (x_{i-1}, x_i)$ such that

$$g(s_i)(x_i - x_{i-1}) = G'(s_i)(x_i - x_{i-1}) = G(x_i) - G(x_{i-1}).$$

If $x_i = x_{i-1}$, let $s_i = x_i$. Then $S = \{s_i, \cdots, s_m\}$ is a sample for P such that

$$\begin{aligned} \sum(g,P,S) &= \sum_{i=1}^m g(s_i)(x_i - x_{i-1}) \\ &= \sum_{i=1}^m G(x_i) - G(x_{i-1}) \\ &= G(x_m) - G(x_0) = G(q) - G(p). \end{aligned}$$

We have shown that for every partition P of $[p,q]$ there is a sample S for P such that

$$\sum(g,P,S) = G(q) - G(p).$$

Let $\{P_n\}$ be a sequence of partitions for $[p,q]$ such that $\{\mu(P_n)\} \to 0$, and for each $n \in \mathbf{Z}^+$ let S_n be a sample for P_n such that

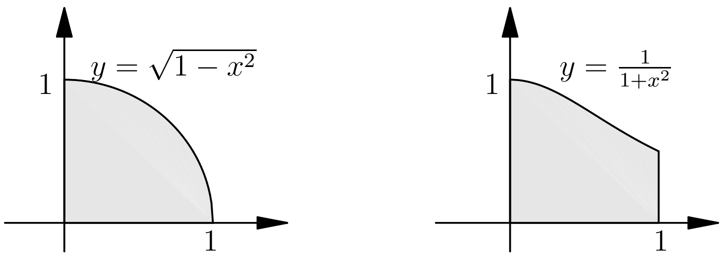
$$\sum(g,P_n,S_n) = G(q) - G(p).$$

Then, since g is integrable on $[q,p]$,

$$\begin{aligned} \int_p^q g &= \lim\{\sum(g,P_n,S_n)\} \\ &= \lim\{G(q) - G(p)\} = G(q) - G(p). \ \| \end{aligned}$$

16.5 Example. The fundamental theorem will allow us to evaluate many integrals easily. For example, we know that $\dfrac{d}{dx}\Big(\arctan(x)\Big) = \dfrac{1}{1+x^2}$. Hence, by the fundamental theorem,

$$\begin{aligned}\int_0^1 \frac{1}{1+x^2}dx &= \arctan(x)\,|_0^1 = \arctan(1) - \arctan(0)\\ &= \frac{\pi}{4} - 0 = \frac{\pi}{4}.\end{aligned}$$



Two sets with the same area

This says that the two sets

$$\{(x,y)\colon 0 \le x \le 1 \text{ and } 0 \le y \le \sqrt{1-x^2}\}$$

and

$$\{(x,y)\colon 0 \le x \le 1 \text{ and } 0 \le y \le \frac{1}{1+x^2}\}$$

have the same area – a rather remarkable result.

16.6 Theorem (Mean value theorem for integrals.) *Let f be a nice function on an interval $[p,q]$, where $p < q$. Then there is a number $c \in (p,q)$ such that*

$$\int_p^q f = f(c)(q-p) \text{ i.e., } f(c) = \frac{1}{q-p}\int_p^q f.$$

Proof: Since f is continuous on $[p,q]$ we can find numbers $r, s \in [p,q]$ such that

$$f(r) \le f(x) \le f(s) \text{ for all } x \in [p,q].$$

By the inequality theorem for integrals

$$\int_p^q f(r) \leq \int_p^q f \leq \int_p^q f(s);$$

(here $f(r)$ and $f(s)$ denote constant functions) i.e.,

$$f(r)(q-p) \leq \int_p^q f \leq f(s)(q-p),$$

i.e.,

$$f(r) \leq \frac{1}{q-p}\int_p^q f \leq f(s).$$

We can now apply the intermediate value property to f on the interval whose endpoints are r and s to get a number c between r and s such that

$$f(c) = \frac{1}{q-p}\int_p^q f.$$

The number c is in the interval (p, q), so we are done. |||

16.7 Corollary. *Let f be a nice function on a closed interval whose endpoints are p and q where $p \neq q$. Then there is a number c between p and q such that*

$$f(c) = \frac{1}{q-p}\int_p^q f.$$

16.8 Exercise. Explain how corollary 16.7 follows from theorem 16.6. (There is nothing to show unless $q < p$)

16.9 Lemma. *Let f be a function such that f is integrable on every subinterval of $[a, b]$. Let $c \in [a, b]$ and let*

$$F(x) = \int_c^x f \text{ for all } x \in [a, b].$$

Then F is continuous on $[a, b]$.

Proof: Let $t \in [a, b]$. I will show that F is continuous at t. Since f is integrable on $[a, b]$ there is a number M such that

$$-M \leq f(x) \leq M \text{ for all } x \in [a, b].$$

By the corollary to the inequality theorem for integrals (8.17), it follows that

$$\Big|\int_s^t f\Big| \leq M|s-t|$$

for all $s,t \in [a,b]$. Thus, for all $s,t \in [a,b]$,

$$0 \leq |F(s)-F(t)| = \Big|\int_c^s f - \int_c^t f\Big| = \Big|\int_t^s f\Big| \leq M|s-t|.$$

Now $\lim\limits_{s\to t} M|s-t| = 0$, so by the squeezing rule for limits of functions, $\lim\limits_{s\to t} |F(s)-F(t)| = 0$. It follows that $\lim_{s\to t} F(s) = F(t)$. |||.

16.10 Theorem (Fundamental theorem of calculus II.) *Let f be a nice function on $[a,b]$, and let $c \in [a,b]$. Let*

$$G(x) = \int_c^x f \text{ for all } x \in [a,b].$$

Then G is an antiderivative for f, i.e.

$$\frac{d}{dx}\int_c^x f = \frac{d}{dx}\int_c^x f(t)\,dt = f(x). \tag{16.11}$$

In particular, every nice function on $[a,b]$ has an antiderivative on $[a,b]$.

Proof: Let

$$G(x) = \int_c^x f \text{ for all } x \in [a,b]$$

and let t be a point in (a,b). Let $\{x_n\}$ be any sequence in $[a,b]\setminus\{t\}$ such that $\{x_n\} \to t$. Then

$$\begin{aligned}\frac{G(x_n)-G(t)}{x_n-t} &= \frac{1}{x_n-t}\Big[\int_c^{x_n} f - \int_c^t f\Big] \\ &= \frac{1}{x_n-t}\int_t^{x_n} f.\end{aligned}$$

By the mean value theorem for integrals, there is a number s_n between x_n and t such that

$$\frac{G(x_n)-G(t)}{x_n-t} = f(s_n).$$

Now

$$0 \leq |s_n - t| \leq |x_n - t| \text{ for all } n$$

and since $\{|x_n - t|\} \to 0$, we have $\{|s_n - t|\} \to 0$, by the squeezing rule for sequences. Since f is continuous, we conclude that $\{f(s_n)\} \to f(t)$; i.e.,

$$\Big\{\frac{G(x_n) - G(t)}{x_n - t}\Big\} \to f(t);$$

i.e.,

$$\lim_{x \to t} \frac{G(x) - G(t)}{x - t} = f(t).$$

This proves that $G'(t) = f(t)$ for $t \in (a, b)$. In addition G is continuous on $[a, b]$ by lemma 16.9. Hence G is an antiderivative for f on $[a, b]$. ‖

Remark Leibnitz's statement of *the fundamental principle of the calculus* was the following:

> Differences and sums are the inverses of one another, that is to say, the sum of the differences of a series is a term of the series, and the difference of the sums of a series is a term of the series; and I enunciate the former thus, $\int dx = x$, and the latter thus, $d \int x = x$[34, page 142].

To see the relation between Leibnitz's formulas and ours, in the equation $d \int x = x$, write $x = ydt$ to get $d \int ydt = ydt$, or $\dfrac{d}{dt} \displaystyle\int ydt = y$. This corresponds to equation (16.11). Equation (16.4) can be written as

$$\int_p^q \frac{dG}{dx} dx = G(q) - G(p).$$

If we cancel the dx's (in the next chapter we will show that this is actually justified!) we get $\int_p^q dG = G(q) - G(p)$. This is not quite the same as $\int dx = x$. However if you choose the origin of coordinates to be $(p, G(p))$, then the two formulas coincide.

To emphasize the inverse-like relation between differentiation and integration, I will restate our formulas for both parts of the the fundamental theorem, ignoring all hypotheses:

$$\frac{d}{dx} \int_c^x f(t) dt = f(x) \text{ and } \int_c^x \frac{d}{dt} f(t) dt = f(x) - f(c).$$

By exploiting the ambiguous notation for indefinite integrals, we can get a form almost identical with Leibniz's:

$$\frac{d}{dx}\int f(x)dx = f(x) \text{ and } \int \frac{d}{dx} f(x)dx = f(x).$$

16.12 Example. Let

$$\begin{aligned} F(x) &= \int_1^x e^{t^2}\, dt, \\ G(x) &= \int_1^{x^3} e^{t^2}\, dt, \\ H(x) &= e^{-x^2} \int_1^x e^{t^2}\, dt. \end{aligned}$$

We will calculate the derivatives of F, G, and H. By the fundamental theorem,

$$F'(x) = e^{x^2}.$$

Now $G(x) = F(x^3)$, so by the chain rule,

$$\begin{aligned} G'(x) &= F'(x^3) \cdot 3x^2 \\ &= e^{(x^3)^2} \cdot 3x^2 = 3x^2 e^{x^6}. \end{aligned}$$

We have $H(x) = e^{-x^2} F(x)$, so by the product rule,

$$\begin{aligned} H'(x) &= e^{-x^2} F'(x) + e^{-x^2}(-2x)F(x) \\ &= e^{-x^2} e^{x^2} + e^{-x^2}(-2x)\int_0^x e^{t^2}\, dt \\ &= 1 - 2xe^{-x^2} \int_0^x e^{t^2}\, dt. \end{aligned}$$

16.13 Exercise. Calculate the derivatives of the following functions. Simplify your answers as much as you can.

a) $F(x) = \displaystyle\int_0^x \frac{1}{\sqrt{1+t^2}}\, dt$

b) $G(x) = \displaystyle\int_0^x \frac{1}{\sqrt{t^2-1}}\, dt$

c) $H(x) = \displaystyle\int_2^x \frac{1}{\sqrt{t^2-1}}\, dt$

d) $K(x) = \int_1^{\sinh(x)} \frac{1}{\sqrt{1+t^2}}\, dt$

e) $L(x) = \int_2^{\cosh(x)} \frac{1}{\sqrt{t^2-1}}\, dt.$

(We defined cosh and sinh in exercise 14.56.) Find simple formulas (not involving any integrals) for K and for L.

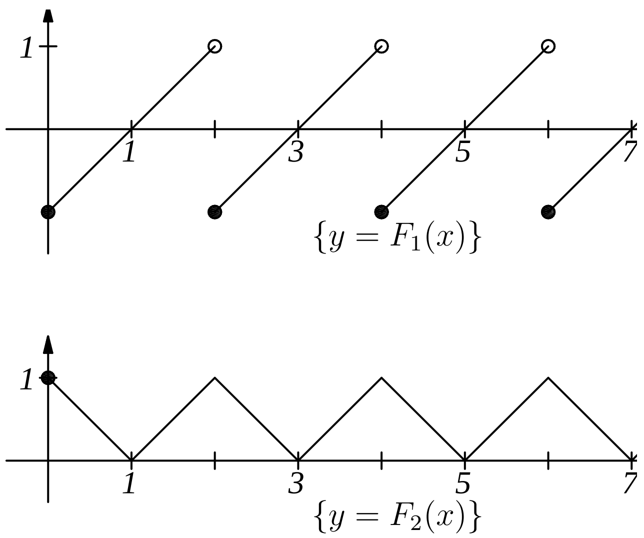
16.14 Exercise. Use the fundamental theorem of calculus to find

a) $\int_0^{\frac{\pi}{4}} \sec^2 x\, dx.$

b) $\int_0^{\frac{1}{\sqrt{2}}} \frac{1}{\sqrt{1-x^2}}\, dx.$

c) $\int_0^1 e^x\, dx.$

16.15 Exercise. Let F_1 and F_2 be the functions whose graphs are shown below:



Let $G_i(x) = \int_0^x F_i(t)\, dt$ for $0 \leq t \leq 8$. Sketch the graphs of G_1 and G_2. Include some discussion about why your answer is correct.

Chapter 17

Antidifferentiation Techniques

17.1 The Antidifferentiation Problem

17.1 Definition ($\int f$ or $\int f(x)dx$.) I am going to use the notation $\int f$ or $\int f(x)dx$ to denote some arbitrary antiderivative for f on an interval that often will not be specified. This is the same notation that I used previously to denote an indefinite integral for f. Although the fundamental theorem of the calculus shows that for nice functions the concepts of "antiderivative" and "indefinite integral" are essentially the same, for arbitrary functions the two concepts do not coincide. For example, let

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0. \end{cases}$$

Then f has an indefinite integral F where

$$F(x) = \begin{cases} \int_0^x 1 \, dx = x & \text{if } x > 0. \\ -\int_0^x 1 \, dx = -x & \text{if } x \leq 0. \end{cases}$$

Thus $F(x) = |x|$. Then F is not an antiderivative for f, since we know that F is not differentiable at 0.

I will always try to make it clear whether $\int f$ represents an antiderivative or an indefinite integral in cases where it makes a difference.

The problem of calculating derivatives is straightforward. By using known formulas and rules, you can easily find the derivative of almost any function you can write down. The problem of calculating antiderivatives is much more complicated. In fact, none of the five functions

$$e^{x^2}, \quad \ln\Big(\ln(x)\Big), \quad \frac{1}{\ln(x)}, \quad \frac{\sin(x)}{x}, \quad \frac{(1-x)^{\frac{3}{5}}}{x^{\frac{12}{5}}} \tag{17.2}$$

have antiderivatives that can be expressed in terms of functions we have studied. (To find a proof of this assertion, see [40, page 37 ff] and [41].) The first two functions in this list are compositions of functions that have simple antiderivatives, the third function is the reciprocal of a function with a simple antiderivative, and each of the last two functions is a product of two functions with simple antiderivatives. (An antiderivative for ln will be calculated in (17.25).) It follows that there is no chain rule or reciprocal rule or product rule for calculating antiderivatives. We will see, however, that the chain rule and the product rule for differentiation do give rise to antidifferentiation formulas.

The five functions

$$e^{\sqrt{x}}, \quad \sin\Big(\ln(x)\Big), \quad \frac{1}{\sin x}, \quad \frac{\ln(x)}{x}, \quad \frac{(1-x)^{\frac{2}{5}}}{x^{\frac{12}{5}}}, \tag{17.3}$$

which look somewhat similar to the functions in (17.2), turn out to have simple antiderivatives, as you will see in (17.42c), (17.22), (17.7), (17.31f) and (17.41). It is often not easy to tell the difference between a function that has a simple antiderivative and a function that does not.

Many simple functions that arise in applied problems do not have simple antiderivatives. The exercises in this chapter have been carefully designed to be non-typical functions whose antiderivatives can be found.

The Maple instructions for finding antiderivatives and integrals are

`int(f(x) , x);` $\quad = \displaystyle\int f(x)dx,$

and

`int(f(x) , x = a..b);` $\quad = \displaystyle\int_a^b f(x)dx.$

I gave the five functions in (17.2) to Maple to antidifferentiate.

The results were:

```
> int( exp(x^2),x);
```

$$-\frac{1}{2}\, I\, \sqrt{\pi}\, \mathrm{erf}(\, I\, x\,)$$

```
> int( ln(ln(x)),x);
```

$$x \ln(\, \ln(\, x\,)\,) + \mathrm{Ei}(\, 1, -\ln(\, x\,)\,)$$

```
> int(1/ln(x),x);
```

$$-\mathrm{Ei}(\, 1, -\ln(\, x\,)\,)$$

```
> int(sin(x)/x,x);
```

$$\mathrm{Si}(\, x\,)$$

```
> int( ((1-x)^(3/5))/ (x^(12/5)),x);
```

$$-\frac{5}{14}\,\frac{2-5\,x+3\,x^2}{x\sqrt[5]{(-1+x)^2\,x^2}} - \int -3/14\,\frac{1}{\sqrt[5]{(-1+x)^2\,x^2}}dx$$

In the first four cases, an answer has been given involving names of functions we have not seen before, (and which we will not see again in this course). The definitions of these functions are:

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t}\, dt, \tag{17.4}$$

$$\mathrm{Ei}(n,x) = \lim_{N\to\infty}\int_1^N \frac{e^{-xt}}{t^n}\, dt, \qquad (n \in \mathbf{Z}^+) \tag{17.5}$$

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\, dt. \tag{17.6}$$

In equation (17.4), we assume that $\frac{\sin(t)}{t} = 1$ when $t = 0$. The function Si is called the *sine integral*. In equation (17.5), $\mathrm{Ei}(n,x)$ makes sense only when x is positive. The definition of $\mathrm{Ei}(n,x)$ for $x < 0$ involves ideas we have not discussed. The function Ei is called the *exponential integral*.

The function erf is called the *error function*. The answer given by Maple for $\int e^{x^2}\, dx$ involves the symbol I. This is Maple's notation for $\sqrt{-1}$. The definition of $\mathrm{erf}(Ix)$ makes no sense in terms of concepts we have studied. However you can use Maple to calculate integrals even if you do not know what the symbols mean. The following instructions find $\int_0^1 \exp(x^2)dx$:

```
> int( exp(x^2),x= 0..1);
```

$$-\frac{1}{2} I \operatorname{erf}(I)\sqrt{\pi}$$

```
> evalf(%);
```

$$1.462651746$$

17.2 Basic Formulas

Every differentiation formula gives rise to an antidifferentiation formula. We review here a list of formulas that you should know. In each case you should verify the formula by differentiating the right side. You should know these formulas backward and forward.

$$\begin{aligned}
\int (f(x))^r f'(x)dx &= \frac{(f(x))^{r+1}}{r+1} \quad (r \neq -1). \\
\int \frac{f'(x)}{f(x)}dx &= \ln(|f(x)|). \\
\int \cos(f(x))f'(x)dx &= \sin(f(x)). \\
\int \sin(f(x))f'(x)dx &= -\cos(f(x)). \\
\int e^{f(x)}f'(x)dx &= e^{f(x)}. \\
\int \sec^2(f(x))f'(x)dx &= \tan(f(x)). \\
\int \csc^2(f(x))f'(x)dx &= -\cot(f(x)). \\
\int \sec(f(x))\tan(f(x))f'(x)dx &= \sec(f(x)). \\
\int \csc(f(x))\cot(f(x))f'(x)dx &= -\csc(f(x)). \\
\int \frac{f'(x)}{1+f^2(x)}dx &= \arctan(f(x)). \\
\int \frac{f'(x)}{\sqrt{1-f^2(x)}}dx &= \arcsin(f(x)).
\end{aligned}$$

17.7 Exercise. Verify that

$$\frac{d}{dx}\Big(\ln(|\sec(f(x))+\tan(f(x))|)\Big) = \sec(f(x))f'(x)$$

and

$$\frac{d}{dx}\Big(\ln(|\csc(f(x)) + \cot(f(x))|)\Big) = -\csc(f(x))f'(x).$$

It follows from the previous exercise that

$$\int \sec(f(x))f'(x)dx = \ln(|\sec(f(x)) + \tan(f(x))|)$$

and

$$\int \csc(f(x))f'(x)dx = -\ln(|\csc(f(x)) + \cot(f(x))|).$$

You should add these two formulas to the list of antiderivatives to be memorized.

17.8 Theorem (Sum rule for antiderivatives) *If f and g are functions that have antiderivatives on some interval $[a, b]$, and if $c \in \mathbf{R}$ then $f + g$, $f - g$ and cf have antiderivatives on $[a, b]$ and*

$$\int (f \pm g) = \int f \pm \int g,$$

and

$$\int cf = c\int f. \tag{17.9}$$

Proof: The meaning of this statement is that if F is an antiderivative for f and G is an antiderivative for G, then $F \pm G$ is an antiderivative for $f \pm g$, and cF is an antiderivative for cf. The warning about the ambiguous notation for indefinite integrals given on page 214 applies also to antiderivatives.

Let F, G be antiderivatives for f and g respectively on $[a, b]$. Then F and G are continuous on $[a, b]$, and

$$F' = f \text{ and } G' = g$$

on (a, b). Hence $F \pm G$ are continuous on $[a, b]$, and

$$(F \pm G)' = F' \pm G' = f \pm g$$

on (a, b), and hence

$$\int (f \pm g) = F \pm G = \int f \pm \int g.$$

Also cF is continuous on $[a, b]$, and

$$(cF)' = cF' = cf$$

on (a, b), so that

$$\int cf = cF = c\int f. \;\|$$

17.10 Example. We will calculate $\int x^2(x^3+1)^3 dx$.
I will try to bring this integral into the form

$$\int (f(x))^r f'(x)dx.$$

It appears reasonable to take $f(x) = (x^3 + 1)$, and then $f'(x) = 3x^2$. The $3x^2$ doesn't quite appear in the integral, but I can get it where I need it by multiplying by a constant, and using (17.9):

$$\int x^2(x^3+1)^3 dx = \frac{1}{3}\int (3x^2)(x^3+1)^3 dx = \frac{1}{3}\frac{(x^3+1)^4}{4} = \frac{(x^3+1)^4}{12}.$$

17.11 Example. We will calculate $\int x(x^3+1)^3 dx$.
This problem is more complicated than the last one. Here I still want to take $f(x) = (x^3 + 1)$, but I cannot get the "$f'(x)$" that I need. I will multiply out $(x^3 + 1)^3$

$$\begin{aligned}\int x(x^3+1)^3 dx &= \int x((x^3)^3 + 3(x^3)^2 + 3(x^3) + 1)dx \\ &= \int (x^{10} + 3x^7 + 3x^4 + x)dx \\ &= \frac{x^{11}}{11} + 3\frac{x^8}{8} + 3\frac{x^5}{5} + \frac{x^2}{2}.\end{aligned}$$

17.12 Example. We will calculate $\int te^{t^2} dt$.

$$\int te^{t^2} dt = \frac{1}{2}\int (2t)e^{t^2} dt.$$

Since $\frac{d}{dt}(t^2) = 2t$ we get

$$\int te^{t^2} dt = \frac{1}{2}e^{t^2}.$$

17.13 Example. We will consider $\int e^{t^2} dt$.
Although this problem looks similar to the one we just did, it can be shown that no function built up from the functions we have studied by algebraic operations is an antiderivative for $\exp(t^2)$. So we will not find the desired antiderivative. (But by the fundamental theorem of the calculus we know that e^{t^2} has an antiderivative.)

17.14 Example. We will calculate $\int \tan$.

$$\int \tan = \int \frac{\sin}{\cos} = -\int \frac{\cos'}{\cos} = -\ln(|\cos|).$$

17.15 Example. We will calculate $\int_0^a \frac{1}{a^2 + x^2}\, dx$.
The integrand $\frac{1}{a^2 + x^2}$ looks enough like $\frac{1}{1 + x^2}$ that I will try to get an arctan from this integral.

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

Now $\frac{d}{dx}\left(\frac{x}{a}\right) = \frac{1}{a}$, so

$$\int \frac{1}{a^2 + x^2} = \frac{1}{a} \int \frac{1}{1 + \left(\frac{x}{a}\right)^2} \frac{d}{dx}\left(\frac{x}{a}\right) dx = \frac{1}{a} \arctan\left(\frac{x}{a}\right).$$

Thus we have found an antiderivative for $\frac{1}{a^2 + x^2}$. Hence

$$\int_0^a \frac{1}{a^2 + x^2}\, dx = \frac{1}{a} \arctan(\frac{x}{a})\Big|_0^a = \frac{1}{a} \arctan(1) = \frac{\pi}{4a}.$$

17.16 Exercise. Find the following antiderivatives:

a) $\int e^x \sin(e^x) dx$.

b) $\int \frac{e^x}{\sin(e^x)} dx$.

c) $\int (3w^4 + w)^2 (12w^3 + 1) dw.$

d) $\int \cos(4x) dx.$

e) $\int \frac{2x}{1 + x^2} dx.$

f) $\int \cot(2x) dx.$

g) $\int \frac{2}{1 + w^2} dw.$

h) $\int \frac{2w}{1 + w^2} dw.$

i) $\int \sin^3(x) dx.$

j) $\int \sin^4(x) dx.$

17.3 Integration by Parts

17.17 Theorem (Integration by parts.) *Let f, g be functions that are continuous on an interval $[a, b]$ and differentiable on (a, b). If $f'g$ has an antiderivative on $[a, b]$, then $g'f$ also has an antiderivative on $[a, b]$ and*

$$\int g'f = fg - \int f'g. \tag{17.18}$$

We call formula (17.18) the formula for integration by parts.

Proof: This theorem is just a restatement of the product rule for differentiation. If f and g are differentiable on (a, b) then the product rule says that

$$(fg)' = f'g + g'f$$

so that

$$g'f = (fg)' - f'g$$

on (a, b). If $\int f'g$ is an antiderivative for $f'g$ on $[a, b]$, then $fg - \int f'g$ is continuous on $[a, b]$, and

$$(fg - \int f'g)' = (fg)' - f'g = g'f$$

on (a, b). We have shown that $fg - \int f'g$ is an antiderivative for $g'f$ on $[a, b]$. ‖

17.19 Example. We will calculate $\int_0^\pi x\sin(3x)dx$. We begin by searching for an antiderivative for $x\sin(3x)$. Let

$$\begin{aligned} f(x) &= x, \\ g'(x) &= \sin(3x), \\ f'(x) &= 1, \\ g(x) &= -\frac{1}{3}\cos(3x). \end{aligned}$$

Then by the formula for integration by parts

$$\begin{aligned} \int x\sin(3x)dx &= \int f(x)g'(x)dx \\ &= f(x)g(x) - \int f'(x)g(x)dx \\ &= -\frac{x}{3}\cos(3x) + \frac{1}{3}\int \cos(3x)dx \\ &= -\frac{x}{3}\cos(3x) + \frac{1}{9}\sin(3x). \end{aligned} \tag{17.20}$$

Hence

$$\begin{aligned} \int_0^\pi x\sin(3x)dx &= \left(-\frac{x}{3}\cos(3x) + \frac{1}{9}\sin(3x)\right)\Big|_0^\pi \\ &= -\frac{\pi}{3}\cos(3\pi) = \frac{\pi}{3}. \end{aligned}$$

Suppose I had tried to find $\int x\sin(3x)$ in the following way: Let

$$\begin{aligned} f(x) &= \sin(3x), \\ g'(x) &= x, \\ f'(x) &= 3\cos(3x), \\ g(x) &= \frac{1}{2}x^2. \end{aligned}$$

Then by the formula for integration by parts

$$\begin{aligned} \int x\sin(3x)dx &= \int f(x)g'(x)dx \\ &= f(x)g(x) - \int f'(x)g(x)dx \\ &= \frac{1}{2}x^2\sin(3x) - \frac{3}{2}\int x^2\cos(3x)dx. \end{aligned} \tag{17.21}$$

In this case the antiderivative $\int x^2\cos(3x)dx$ looks more complicated than the one I started out with. When you use integration by parts, it is not always clear what you should take for f and for g'. If you find things are starting to look more complicated rather than less complicated, you might try another choice for f and g'.

Integration by parts is used to evaluate antiderivatives of the forms $\displaystyle\int x^n\sin(ax)dx$, $\displaystyle\int x^n\cos(ax)dx$, and $\displaystyle\int x^n e^x dx$ when n is a positive integer. These can be reduced to antiderivatives of the forms $\displaystyle\int x^{n-1}\sin(ax)dx$, $\displaystyle\int x^{n-1}\cos(ax)dx$, and $\displaystyle\int x^{n-1}e^x dx$, so by applying the process n times we get the power of x down to x^0, which gives us antiderivatives we can easily find.

17.22 Example. We will calculate $\int\sin(\ln(x))$.
Let

$$\begin{aligned} f(x) &= \sin(\ln(x)), \\ g'(x) &= 1, \\ g(x) &= x, \\ f'(x) &= \frac{\cos(\ln(x))}{x}. \end{aligned}$$

Then

$$\begin{aligned} \int\sin(\ln(x))dx &= \int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx \\ &= x\sin(\ln(x)) - \int\cos(\ln(x))dx \end{aligned} \tag{17.23}$$

We will now use integration by parts to find an antiderivative for $\cos(\ln(x))$.
Let

$$F(x) = \cos(\ln(x)),$$

$$\begin{aligned} G'(x) &= 1, \\ G(x) &= x, \\ F'(x) &= -\frac{\sin(\ln(x))}{x}. \end{aligned}$$

Then

$$\begin{aligned} \int \cos(\ln(x))dx &= \int F(x)G'(x) = F(x)G(x) - \int F'(x)G(x)dx \\ &= x\cos(\ln(x)) + \int \sin(\ln(x))dx \end{aligned} \tag{17.24}$$

From equations (17.23) and (17.24) we see that

$$\int \sin(\ln(x))dx = x\sin(\ln(x)) - \left(x\cos(\ln(x)) + \int \sin(\ln(x))\right).$$

Thus

$$2\int \sin(\ln(x))dx = x\sin(\ln(x)) - x\cos(\ln(x)),$$

and

$$\int \sin(\ln(x))dx = \frac{x}{2}(\sin(\ln(x)) - \cos(\ln(x))).$$

17.25 Example. We will calculate $\int \ln(t)dt$. Let

$$\begin{aligned} f(t) &= \ln(t), \\ g'(t) &= 1, \\ g(t) &= t, \\ f'(t) &= \frac{1}{t}. \end{aligned}$$

Then

$$\begin{aligned} \int \ln(t)dt &= \int f(t)g'(t)dt = f(t)g(t) - \int f'(t)g(t)dt \\ &= t\ln(t) - \int 1dt \\ &= t\ln(t) - t. \end{aligned}$$

17.26 Theorem (Antiderivative of inverse functions.) *Let I and J be intervals in $\mathbf{R}$, and let $f\colon I \to J$ be a continuous function such that $f'(x)$ is defined and non-zero for all x in the interior of I. Suppose that $g\colon J \to I$ is* *the inverse function for f, and that F is an antiderivative for f. Then an antiderivative for g on J is given by*

$$\int g(x)\,dx = xg(x) - (F \circ g)(x). \tag{17.27}$$

Proof: Let $h(x) = x$. Then $h'(x) = 1$, and

$$\begin{aligned}\int g(x)\,dx &= \int g(x)h'(x)dx = g(x)h(x) - \int g'(x)h(x)dx \\ &= xg(x) - \int xg'(x)dx \end{aligned} \tag{17.28}$$

Now $F' = f$ and $f \circ g(x) = x$ for all x in J, so

$$(F \circ g)'(x) = F'\Big(g(x)\Big) \cdot g'(x) = f\Big(g(x)\Big) \cdot g'(x) = xg'(x)$$

and it follows from (17.28) that

$$\begin{aligned}\int g(x)\,dx &= xg(x) - \int (F \circ g)'(x)\,dx \\ &= xg(x) - (F \circ g)(x). \;\|\end{aligned}$$

Remark: It follows from the proof of the previous theorem that if you know an antiderivative for a function f, then you can find an antiderivative for the inverse function g by integration by parts. This is what you should remember about the theorem. The formula (17.27) is not very memorable.

17.29 Example. In the previous theorem, if we take

$$f(x) = e^x, \quad F(x) = e^x, \quad g(x) = \ln(x),$$

then we get

$$\int \ln(x)\,dx = x\ln(x) - e^{\ln(x)} = x\ln(x) - x.$$

This agrees with the result obtained in example 17.25.

17.30 Exercise. What is wrong with the following argument? Let

$$\begin{aligned} f(x) &= \frac{1}{x}, \\ g'(x) &= 1, \\ f'(x) &= -\frac{1}{x^2}, \\ g(x) &= x. \end{aligned}$$

Then

$$\begin{aligned}\int \frac{1}{x}dx &= \int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx \\ &= 1 + \int \frac{1}{x}dx.\end{aligned}$$

If we subtract $\int \frac{1}{x}dx$ from both sides we obtain

$$0 = 1.$$

17.31 Exercise. Calculate the following antiderivatives:

a) $\int xe^x dx$.

b) $\int e^x \sin(x)dx$. (Integrate by parts more than once.)

c) $\int \arctan(u)du$.

d) $\int \frac{x}{\sqrt{4-x^2}}dx$.

e) $\int x\sqrt{4-x^2}dx$.

f) $\int x^r \ln(|x|)dx$, where $r \in \mathbf{R}$. Have you considered the case where $r = -1$?

g) $\int x^2 \cos(2x)dx$.

17.4 Integration by Substitution

We will now use the chain rule to find some antiderivatives. Let g be a real valued function that is continuous on some interval J and differentiable on the interior of J, and let f be a function such that f has an antiderivative F on some interval K. We will suppose that $g(J) \subset K$ and $g\big(\text{interior}(J)\big) \subset \text{interior}(K)$. It then follows that $F \circ g$ is continuous on J and differentiable on interior(J), and

$$(F \circ g)'(t) = (F' \circ g)(t)g'(t) = (f \circ g)(t)g'(t) \tag{17.32}$$

for all t in the interior of J; i.e., $F \circ g$ is an antiderivative for $(f \circ g)g'$ on J. Thus

$$\int f\Big(g(t)\Big)g'(t)dt = F\Big(g(t)\Big) \text{ where } F(u) = \int f(u)du. \tag{17.33}$$

There is a standard ritual for using (17.33) to find $\int f\Big(g(t)\Big)g'(t)dt$ when an antiderivative F can be found for f. We write:

Let $u = g(t)$. Then $du = g'(t)dt$ (or $du = \frac{du}{dt}dt$), so

$$\int f\Big(g(t)\Big)g'(t)dt = \int f(u)du = F(u) = F\Big(g(t)\Big). \tag{17.34}$$

In the first equality of (17.34) we replace $g(t)$ by u and $g'(t)dt$ by du, and in the last step we replace u by $g(t)$. Since we have never assigned any meaning to "*du*" or "*dt*", we should think of (17.34) just as a mnemonic device for remembering (17.33).

17.35 Example. Find $\int \frac{\sin(\sqrt{x})}{\sqrt{x}}dx$.

Let $u = \sqrt{x}$. Then $du = \frac{1}{2\sqrt{x}}dx$, so

$$\begin{aligned}
\int \frac{\sin(\sqrt{x})}{\sqrt{x}}dx &= 2\int \sin(\sqrt{x})\frac{1}{2\sqrt{x}}dx \\
&= 2\int \sin(u)du = -2\cos(u) \\
&= -2\cos(\sqrt{x}). \ \|\!|
\end{aligned}$$

Suppose we want to find $\int \sin(\sqrt{x})dx$. If we had a $\sqrt{x}$ in the denominator, this would be a simple problem. (In fact we just considered this problem in the previous example.) We will now discuss a method of introducing the missing $\sqrt{x}$.

Suppose g is a function on an interval J such that $g'(t)$ is never zero on the interior of J, and suppose that h is an inverse function for g. Then

$$\Big(h\Big(g(x)\Big) = x\Big) \implies \Big(h'\Big(g(x)\Big) \cdot g'(x) = 1\Big)$$

for all x in the interior of J, so

$$\int f\Big(g(x)\Big)dx = \int f\Big(g(x)\Big) \cdot h'\Big(g(x)\Big) \cdot g'(x)dx.$$

We now apply the ritual (17.34): Let $u = g(x)$. Then $du = g'(x)dx$, so

$$\begin{aligned} \int f\Big(g(x)\Big)dx &= \int f\Big(g(x)\Big)h'\Big(g(x)\Big) \cdot g'(x)dx \\ &= \int f(u)h'(u)du. \end{aligned}$$

If we can find an antiderivative H for fh', then

$$\int f(u)h'(u)du = H(u) = H\Big(g(x)\Big).$$

We have shown that if h is an inverse function for g, then

$$\int f\Big(g(x)\Big)dx = H\Big(g(x)\Big) \text{ where } H(u) = \int f(u)h'(u)du \tag{17.36}$$

There is a ritual associated with this result also. To find $\int f\Big(g(x)\Big)dx$:

Let $u = g(x)$. Then $x = h(u)$ so $dx = h'(u)du$.

Hence

$$\int f\Big(g(x)\Big)dx = \int f(u)h'(u)du = H(u) = H\Big(g(x)\Big). \tag{17.37}$$

17.38 Example. To find $\int \sin(\sqrt{x})dx$.

Let $u = \sqrt{x}$. Then $x = u^2$ so $dx = 2u\ du$.

Thus

$$\int \sin(\sqrt{x})dx = \int \sin(u) \cdot 2u\ du = 2\int u\sin(u)du.$$

We can now use integration by parts to find $\int u\sin(u)du$. Let

$$\begin{aligned} f(u) = u, \qquad g'(u) &= \sin(u), \\ f'(u) = 1, \qquad g(u) &= -\cos(u). \end{aligned}$$

Then

$$\begin{aligned}\int u\sin(u)du &= \int f(u)g'(u)du \\ &= f(u)g(u) - \int f'(u)g(u)du \\ &= -u\cos(u) + \int \cos(u)du \\ &= -u\cos(u) + \sin(u).\end{aligned}$$

Hence

$$\begin{aligned}\int \sin(\sqrt{x})dx &= 2\int u\sin u\ du \\ &= -2u\cos(u) + 2\sin(u) \\ &= -2\sqrt{x}\cos(\sqrt{x}) + 2\sin(\sqrt{x}).\end{aligned}$$

17.39 Example. To find $\int \frac{1}{e^x + e^{-x}}dx$.

Let $u = e^x$. Then $x = \ln(u)$ so $dx = \frac{1}{u}du$.

$$\begin{aligned}\int \frac{1}{e^x + e^{-x}}dx &= \int \frac{1}{(u+\frac{1}{u})}\cdot\frac{1}{u}du = \int \frac{1}{u^2+1}du \\ &= \arctan(u) = \arctan(e^x).\end{aligned}$$

17.40 Example. To find $\int t\sqrt{t+1}\ dt$.

Let $u = t+1$. Then $t = u-1$ so $dt = du$.

Hence

$$\begin{aligned}\int t\sqrt{t+1}\ dt &= \int (u-1)\sqrt{u}\ du = \int u^{3/2} - u^{1/2}du \\ &= \frac{2}{5}u^{5/2} - \frac{2}{3}u^{3/2} = \frac{2}{5}(t+1)^{5/2} - \frac{2}{3}(t+1)^{3/2}.\end{aligned}$$

17.41 Example. To find $\int \frac{(1-x)^{\frac{2}{5}}}{x^{\frac{12}{5}}} dx$.

$$\int \frac{(1-x)^{\frac{2}{5}}}{x^{\frac{12}{5}}} dx = \int \left(\frac{1-x}{x}\right)^{\frac{2}{5}} \cdot \frac{1}{x^2} dx.$$

Let $u = \frac{1-x}{x} = \frac{1}{x} - 1$. Then $du = -\frac{1}{x^2} dx$, and

$$\int \frac{(1-x)^{\frac{2}{5}}}{x^{\frac{12}{5}}} dx = -\int u^{\frac{2}{5}} du = -\frac{5}{7} u^{\frac{7}{5}} = -\frac{5}{7}\left(\frac{1-x}{x}\right)^{\frac{7}{5}}.$$

Thus

$$\int \frac{(1-x)^{\frac{2}{5}}}{x^{\frac{12}{5}}} dx = -\frac{5}{7}\left(\frac{1-x}{x}\right)^{\frac{7}{5}}.$$

17.42 Exercise. Find the following antiderivatives:

a) $\int x^2 \sin(x^3) dx$.

b) $\int \frac{e^x}{1+e^x} dx$.

c) $\int e^{\sqrt{x}} dx$.

d) $\int \frac{\ln(3x)}{x} dx$.

e) $\int 2^x \, dx$.

f) $\int \frac{e^{2x}+e^{3x}}{e^{4x}} \, dx$.

g) $\int x(1+\sqrt[3]{x}) dx$.

17.5 Trigonometric Substitution

Integrals of the form $\int F(\sqrt{a^2+x^2})dx$ and $\int F(\sqrt{a^2-x^2})dx$ often arise in applications. There is a special trick for dealing with such integrals. Since

$$x = a\tan(\arctan(\frac{x}{a})) \text{ for all } x \in \mathbf{R},$$

we can write

$$\int F(\sqrt{a^2+x^2})dx = \int F(\sqrt{a^2+\Big(a\tan(\arctan(\frac{x}{a}))\Big)^2})dx.$$

If we now make the substitution

$$u = \arctan(\frac{x}{a}) \text{ or } x = a\tan(u), \qquad \Big(u \in (-\frac{\pi}{2}, \frac{\pi}{2})\Big)$$

then we find $dx = a\sec^2(u)du$, or

$$\int F(\sqrt{a^2+x^2})dx = \int F(\sqrt{a^2+\Big(a\tan(u)\Big)^2})a\sec^2 u\, du.$$

Now

$$a^2 + \Big(a\tan(u)\Big)^2 = a^2\Big(1+\tan^2(u)\Big) = a^2\sec^2(u)$$

so

$$\sqrt{a^2+\Big(a\tan(u)\Big)^2} = a\sec(u).$$

(Since $u \in \Big(-\frac{\pi}{2}, \frac{\pi}{2}\Big)$ we have $\sec(u) > 0$ and the square root is positive.) Thus

$$\int F(\sqrt{a^2+x^2})dx = a\int F\Big(a\sec(u)\Big)\cdot\sec^2(u)du.$$

Often this last antiderivative can be found. If

$$a\int F\Big(a\sec(u)\Big)\cdot\sec^2(u)du = H(u),$$

then by the ritual (17.37)

$$\int F(\sqrt{a^2+x^2})dx = a\int F\Big(a\sec(u)\Big)\cdot\sec^2(u)du = H(u) = H(\arctan(\frac{x}{a})).$$

The ritual to apply when using this method for finding $\int F(\sqrt{a^2+x^2})dx$ is:

Let $x = a\tan(u)$. Then $dx = a\sec^2(u)du$, and

$$\sqrt{a^2+x^2} = \sqrt{a^2+a^2\tan^2(u)} = \sqrt{a^2\sec^2(u)} = a\sec(u),$$

so

$$\int F(\sqrt{a^2+x^2})dx = a\int F\Big(a\sec(u)\Big)\sec^2(u)du = H(u) = H(\arctan(\frac{x}{a})).$$

There is a similar ritual for integrals of the form $\int F(\sqrt{a^2-x^2})dx$ (Here we will just describe the ritual).

Let $x = a\sin(u)$. Then $dx = a\cos(u)du$ and

$$\sqrt{a^2-x^2} = \sqrt{a^2-a^2\sin(u)} = \sqrt{a^2\cos^2(u)} = a\cos(u) \tag{17.43}$$

so

$$\int F(\sqrt{a^2-x^2})dx = a\int F\Big(a\cos(u)\Big)\cdot\cos(u)du = H(u) = H(\arcsin(\frac{x}{a})).$$

Observe that in equation (17.43) we are assuming that $u = \arcsin(\frac{x}{a})$, so $u \in (-\frac{\pi}{2}, \frac{\pi}{2})$, so $\cos(u) \geq 0$, and the sign of the square root is correct.

17.44 Example. Find $\int\sqrt{4+x^2}dx$.

Let $x = 2\tan\theta$. Then $dx = 2\sec^2\theta\, d\theta$, and

$$\sqrt{4+x^2} = \sqrt{4(1+\tan^2\theta)} = 2\sqrt{\sec^2(\theta)} = 2\sec(\theta). \tag{17.45}$$

Thus

$$\int\sqrt{4+x^2}\,dx = 2^2\int\sec\theta\cdot\sec^2\theta\,d\theta = 4\int\sec^3(\theta)\,d\theta.$$

To find $\int\sec^3(\theta)d\theta$, I will integrate by parts. Let

$$\begin{array}{ll} f(\theta) = \sec(\theta), & g'(\theta) = \sec^2(\theta), \\ f'(\theta) = \sec(\theta)\tan(\theta), & g(\theta) = \tan(\theta). \end{array}$$

Hence,

$$\begin{aligned}\int \sec^3(\theta)d\theta &= \int f(\theta)g'(\theta)d\theta \\ &= f(\theta)g(\theta) - \int f'(\theta)g(\theta)d\theta \\ &= \sec(\theta)\tan(\theta) - \int \sec(\theta)\tan^2(\theta)d\theta \\ &= \sec(\theta)\tan(\theta) - \int \sec(\theta)(\sec^2(\theta)-1)d\theta \\ &= \sec(\theta)\tan(\theta) - \int \sec^3(\theta)d\theta + \int \sec(\theta)d\theta.\end{aligned}$$

Hence

$$\begin{aligned}2\int \sec^3(\theta)d\theta &= \sec(\theta)\tan(\theta) + \int \sec(\theta)d\theta \\ &= \sec(\theta)\tan(\theta) + \ln(|\sec(\theta)+\tan(\theta)|);\end{aligned}$$

i.e.,

$$\int \sec^3(\theta)d\theta = \frac{1}{2}\Bigg(\sec(\theta)\tan(\theta) + \ln\Big(|\sec(\theta)+\tan(\theta)|\Big)\Bigg). \tag{17.46}$$

Hence

$$\begin{aligned}\int \sqrt{4+x^2}dx &= 4\int \sec^3(\theta)d\theta \\ &= 2\Bigg(\sec(\theta)\tan(\theta) + \ln\Big(|\sec(\theta)+\tan(\theta)|\Big)\Bigg).\end{aligned}$$

By (17.45) we have $\tan(\theta) = \dfrac{x}{2}$ and $\sec(\theta) = \dfrac{1}{2}\sqrt{4+x^2}$. Thus

$$\begin{aligned}\int \sqrt{4+x^2}dx &= 2\Bigg(\frac{1}{2}\sqrt{4+x^2}\cdot\frac{x}{2} + \ln\Big(\Big|\frac{\sqrt{4+x^2}}{2}+\frac{x}{2}\Big|\Big)\Bigg) \\ &= \frac{x\sqrt{4+x^2}}{2} + 2\ln\Big(\Big|\frac{\sqrt{4+x^2}+x}{2}\Big|\Big).\end{aligned}$$

17.47 Example. In the process of working out the last example we found $\int \sec^3(\theta)d\theta$ using integration by parts. Here is a different tricky way of finding the same integral [32].

$$\begin{aligned}
\int \sec^3(\theta)d\theta &= \frac{1}{2}\int(\sec^3(\theta)+\sec^3(\theta))d\theta \\
&= \frac{1}{2}\int(\sec(\theta)(1+\tan^2(\theta))+\sec^3(\theta))d\theta \\
&= \frac{1}{2}\int(\sec(\theta)+\big((\sec(\theta)\tan(\theta))\cdot\tan(\theta)+\sec(\theta)\cdot\sec^2(\theta)\big))d\theta \\
&= \frac{1}{2}\int(\sec(\theta)+\frac{d}{d\theta}(\sec(\theta)\tan(\theta)))d\theta \\
&= \frac{1}{2}(\ln(|\sec(\theta)+\tan(\theta)|)+\sec(\theta)\tan(\theta)).
\end{aligned}$$

17.48 Example. Find $\displaystyle\int \frac{1}{\sqrt{a^2-x^2}}dx.$

Let $x = a\sin(\theta)$. Then $dx = a\cos(\theta)d\theta$ and

$$\sqrt{a^2-x^2} = \sqrt{a^2-a^2\sin^2\theta} = a\sqrt{\cos^2\theta} = a\cos\theta.$$

Thus

$$\begin{aligned}
\int \frac{1}{\sqrt{a^2-x^2}}dx &= \int \frac{a\cos(\theta)}{a\cos(\theta)}d\theta = \int 1\ d\theta \\
&= \theta = \arcsin(\frac{x}{a}).
\end{aligned}$$

17.49 Exercise. Find the following antiderivatives:

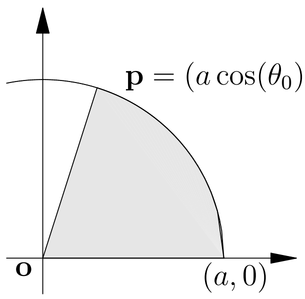
a) $\displaystyle\int \sqrt{a^2-x^2}\ dx$

b) $\displaystyle\int \frac{1}{\sqrt{a^2+x^2}}\ dx$

c) $\displaystyle\int \frac{x}{\sqrt{a^2-x^2}}\ dx$

d) $\displaystyle\int x\sqrt{a^2+x^2}\ dx$

17.50 Example (Area of a circular sector) Let a be a positive number, and let θ_0 be a number in $[0, \frac{\pi}{2})$. Let $\mathbf{o} = (0,0)$, and let $\mathbf{p} = (a\cos(\theta_0), a\sin(\theta_0))$. Let $T(a,\theta_0)$ be the circular sector bounded by the positive x-axis, the segment $[\mathbf{op}]$, and the circle $\{x^2 + y^2 = a^2\}$.



$T(a,\theta_0)$ is shaded region

The equation for $[\mathbf{op}]$ is

$$y = \frac{a\sin(\theta_0)}{a\cos(\theta_0)}x = x\tan(\theta_0),$$

and the equation for the upper semicircle is

$$y = \sqrt{a^2 - x^2}.$$

Hence

$$\text{area}(T(a,\theta_0)) = A_0^a(f),$$

where

$$f(x) = \begin{cases} x\tan(\theta_0) & \text{if } 0 \le x \le a\cos(\theta_0), \\ \sqrt{a^2 - x^2} & \text{if } a\cos(\theta_0) \le x \le a. \end{cases}$$

i.e.

$$\text{area}(T(a,\theta_0)) = \int_0^{a\cos(\theta_0)} x\tan(\theta_0)dx + \int_{a\cos(\theta_0)}^a \sqrt{a^2 - x^2}\, dx.$$

In exercise 17.49.a you showed that

$$\int \sqrt{a^2 - x^2} = \frac{1}{2}a^2 \arcsin\left(\frac{x}{a}\right) + \frac{1}{2}x\sqrt{a^2 - x^2},$$

so

$$\begin{aligned}
\text{area}(T(a,\theta_0)) &= \tan(\theta_0)\,\frac{x^2}{2}\bigg|_0^{a\cos(\theta_0)} + \left(\frac{1}{2}a^2\arcsin(\frac{x}{a}) + \frac{1}{2}x\sqrt{a^2-x^2}\right)\bigg|_{a\cos(\theta_0)}^{a} \\
&= \frac{1}{2}\tan(\theta_0)a^2\cos^2(\theta_0) + \frac{1}{2}a^2\arcsin(1) \\
&\qquad -\frac{1}{2}a^2\arcsin(\cos(\theta_0)) - \frac{1}{2}a\cos(\theta_0)\sqrt{a^2-a^2\cos^2(\theta_0)} \\
&= \frac{a^2}{2}\sin(\theta_0)\cos(\theta_0) + \frac{\pi a^2}{4} \\
&\qquad -\frac{a^2}{2}\arcsin(\sin(\frac{\pi}{2}-\theta_0)) - \frac{a^2}{2}\cos(\theta_0)\sqrt{1-\cos^2(\theta_0)} \\
&= \frac{a^2}{2}\sin(\theta_0)\cos(\theta_0) + \frac{\pi a^2}{4} - \frac{a^2}{2}(\frac{\pi}{2}-\theta_0) - \frac{a^2}{2}\sin(\theta_0)\cos(\theta_0) \\
&= \frac{1}{2}a^2\theta_0.
\end{aligned}$$

By using symmetry arguments, you can show that this formula actually holds for $0 \le \theta_0 \le 2\pi$.

17.6 Substitution in Integrals

Let f be a nice function on an interval $[a,b]$. Then if F is any antiderivative for f, we have

$$\int_a^b f = F\,|_a^b = F(b) - F(a),$$

by the fundamental theorem of calculus. We saw in (17.32) that under suitable hypotheses on g, $F \circ g$ is an antiderivative for $(f \circ g)g'$. Hence

$$\int_a^b f\big(g(t)\big)g'(t)dt = F \circ g\,|_a^b = F\Big|_{g(a)}^{g(b)} = \int_{g(a)}^{g(b)} f(u)du.$$

Hence we can find $\displaystyle\int_a^b f\big(g(t)\big)g'(t)dt$ by the following ritual:

Let $u = g(t)$. When $t = a$ then $u = g(a)$ and when $t = b$ then $u = g(b)$. Also $du = g'(t)dt$. Hence

$$\int_a^b f\big(g(t)\big)g'(t)dt = \int_{g(a)}^{g(b)} f(u)du = F(u)\Big|_{g(a)}^{g(b)}.$$

17.51 Example. To find $\displaystyle\int_{\pi^2}^{4\pi^2} \frac{\sin(\sqrt{x})}{\sqrt{x}}dx$.

Let $u = \sqrt{x}$. When $x = \pi^2$, then $u = \pi$, and when $x = 4\pi^2$, then $u = 2\pi$. Also $du = \dfrac{1}{2\sqrt{x}}dx$, so

$$\begin{aligned}\int_{\pi^2}^{4\pi^2} \frac{\sin(\sqrt{x})}{\sqrt{x}}dx &= 2\int_{\pi}^{2\pi} \sin(u)du = -2\cos u\Big|_{\pi}^{2\pi} \\ &= -2\Big(\cos(2\pi) - \cos(\pi)\Big) = -2(1+1) = -4.\end{aligned}$$

We saw in (17.36) that if h is an inverse function for g, then an antiderivative for $f \circ g$ is $H \circ g$, where H is an antiderivative for $f \cdot h'$. Thus

$$\int_a^b f\Big(g(t)\Big)dt = H \circ g \;|_a^b = H\Big|_{g(a)}^{g(b)}.$$

The ritual for finding $\displaystyle\int_a^b f\Big(g(t)\Big)dt$ in this case is:

Let $u = g(t)$. Then $t = h(u)$ and $dt = h'(u)du$. When $t = a$ then $u = g(a)$, and when $t = b$ then $u = g(b)$. Thus

$$\int_a^b f\Big(g(t)\Big)dt = \int_{g(a)}^{g(b)} f(u)h'(u)du = H(u)\Big|_{g(a)}^{g(b)}$$

where H is an antiderivative for fh'.

17.52 Example. To find $\displaystyle\int_0^{\ln(\sqrt{3})} \frac{1}{e^x + e^{-x}}dx$.

Let $u = e^x$. When $x = 0$ then $u = 1$, and when $x = \ln(\sqrt{3})$ then $u = \sqrt{3}$. Also $x = \ln(u)$, so $dx = \dfrac{1}{u}du$.

$$\begin{aligned}\int_0^{\ln(\sqrt{3})} \frac{1}{e^x + e^{-x}}dx &= \int_1^{\sqrt{3}} \frac{1}{(u + \frac{1}{u})} \cdot \frac{1}{u}du \\ &= \int_1^{\sqrt{3}} \frac{1}{u^2+1}du = \arctan(u) \;|_1^{\sqrt{3}} \\ &= \arctan(\sqrt{3}) - \arctan(1) \\ &= \frac{\pi}{3} - \frac{\pi}{4} = \frac{\pi}{12}.\end{aligned}$$

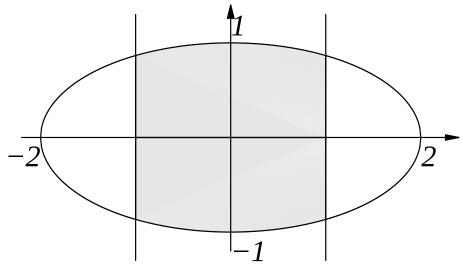
17.53 Exercise. Find the following integrals:

a) $\int_0^1 x^2(x^3+1)^3 dx.$

b) $\int_0^{3/2} \frac{1}{\sqrt{9-x^2}} dx.$

c) $\int_0^1 x\sqrt{1-x}\, dx.$

17.54 Exercise. Find the area of the shaded region, bounded by the ellipse $\frac{x^2}{4}+y^2=1$ and the lines $x=\pm 1$.



17.55 Example. In practice I would find many of the antiderivatives and integrals discussed in this chapter by computer. For example, using Maple, I would find

```
> int(sqrt(a^2+x^2),x);
```

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

```
> int(sin(sqrt(x)),x=0..Pi^2);
```

$$2\,\pi$$

```
> int(sqrt(4 - x^2),x=-1..1);
```

$$\sqrt{3}+\frac{2}{3}\,\pi$$

```
> int( (sec(x))^3,x);
```

$$\frac{1}{2}\,\frac{\sin(\,x\,)}{\cos(\,x\,)^2}+\frac{1}{2}\ln(\,\sec(\,x\,)+\tan(\,x\,)\,)$$

```
> int(exp(a*x)*cos(b*x),x);
```

$$\frac{a\,e^{(a\,x)}\cos(\,b\,x\,)}{a^2+b^2}+\frac{b\,e^{(a\,x)}\sin(\,b\,x\,)}{a^2+b^2}$$

17.7 Rational Functions

In this section we present a few rules for finding antiderivatives of simple rational functions.

To antidifferentiate $\dfrac{P(x)}{(x-c)^n}$ where P is a polynomial, make the substitution $u = x - c$.

17.56 Example. To find $\displaystyle\int \frac{x^2+1}{(x-2)^2}dx$.

Let $u = x - 2$. Then $x = 2 + u$ so $dx = du$, and

$$\begin{aligned}
\int \frac{(x^2+1)}{(x-2)^2}dx &= \int \frac{(2+u)^2+1}{u^2}du \\
&= \int \frac{u^2+4u+5}{u^2}du \\
&= \int 1+\frac{4}{u}+\frac{5}{u^2}du \\
&= u+4\ln|u|-\frac{5}{u} \\
&= (x-2)+4\ln(|x-2|)-\frac{5}{(x-2)}.
\end{aligned}$$

To find $\displaystyle\int \frac{R(x)}{(x-a)(x-b)}dx$ where $a \neq b$ and R is a polynomial of degree less than 2.

We will find numbers A and B such that

$$\frac{R(x)}{(x-a)(x-b)} = \frac{A}{(x-a)}+\frac{B}{(x-b)}. \tag{17.57}$$

Suppose (17.57) were valid. If we multiply both sides by $(x - a)$ we get

$$\frac{R(x)}{(x-b)} = A+\frac{B(x-a)}{x-b}.$$

Now take the limit as x goes to a to get

$$\frac{R(a)}{a-b} = A.$$

The reason I took a limit here, instead of saying "now for $x = a$ we get $\cdots$" is that a is not in the domain of the function we are considering. Similarly

$$\frac{R(x)}{x-a} = \frac{A(x-b)}{x-a} + B,$$

and if we take the limit as x goes to b, we get

$$\frac{R(b)}{b-a} = B.$$

Thus,

$$\frac{R(x)}{(x-a)(x-b)} = \frac{1}{a-b}\left[\frac{R(a)}{x-a} - \frac{R(b)}{x-b}\right]. \qquad (17.58)$$

I have now shown that if there are numbers A and B such that (17.57) holds, then (17.58) holds. Since I have not shown that such numbers exist, I will verify directly that (17.58) is valid. Write $R(x) = px + q$. Then

$$\begin{aligned}
\frac{1}{a-b}\left[\frac{R(a)}{x-a} - \frac{R(b)}{x-b}\right] &= \frac{1}{a-b}\left[\frac{pa+q}{x-a} - \frac{pb+q}{x-b}\right] \\
&= \frac{1}{(a-b)}\left[\frac{(pa+q)(x-b) - (pb+q)(x-a)}{(x-a)(x-b)}\right] \\
&= \frac{1}{(a-b)}\left[\frac{x(pa-pb) - q(b-a)}{(x-a)(x-b)}\right] \\
&= \frac{1}{(a-b)}\frac{(a-b)(px+q)}{(x-a)(x-b)} \\
&= \frac{px+q}{(x-a)(x-b)} = \frac{R(x)}{(x-a)(x-b)}. \quad \|
\end{aligned}$$

17.59 Example. To find $\displaystyle\int \frac{x+1}{(x+2)(x+3)}dx.$

Let $\displaystyle\frac{x+1}{(x+2)(x+3)} = \frac{A}{x+2} + \frac{B}{x+3}.$

Then

$$\frac{x+1}{x+3} = A + \frac{x+2}{x+3}B,$$

so

$$A = \frac{-2+1}{-2+3} = -1,$$

and

$$\frac{x+1}{x+2} = A\frac{(x+3)}{x+2} + B,$$

so

$$B = \frac{-3+1}{-3+2} = 2.$$

Hence

$$\begin{aligned}\int \frac{x+1}{(x+2)(x+3)}dx &= \int \frac{-1}{x+2} + \frac{2}{x+3}dx \\ &= -\ln(|x+2|) + 2\ln(|x+3|).\end{aligned}$$

In this example I did not use formula (17.58), because I find it easier to remember the procedure than the general formula. I do not need to check my answer, because my proof of (17.58) shows that the procedure always works. (In practice, I usually do check the result because I am likely to make an arithmetic error.)

To find $\displaystyle\int \frac{R(x)}{x^2+ax+b}dx$ where R is a polynomial of degree < 2, and x^2+ax+b does not factor as a product of two first degree polynomials.

Complete the square to write

$$x^2 + ax + b = (x-m)^2 + k.$$

Then $k > 0$, since if $k = 0$ then we have factored $x^2 + ax + b$, and if $k < 0$ we can write $k = -n^2$, and then

$$(x-m)^2 + k = (x-m)^2 - n^2 = \Big((x-m) - n\Big)\Big((x-m) + n\Big)$$

and again we get a factorization of $x^2 + ax + b$. Since $k > 0$, we can write $k = q^2$ for some $q \in \mathbf{R}$, and

$$x^2 + ax + b = (x-m)^2 + q^2.$$

Now

$$\int \frac{R(x)}{x^2+ax+b}dx = \int \frac{R(x)}{(x-m)^2+q^2}dx.$$

Make the substitution $u = x - m$ to get an antiderivative of the form

$$\begin{aligned}\int \frac{Au+B}{u^2+q^2}du &= \frac{A}{2}\int \frac{2u}{u^2+q^2}du + B\int \frac{1}{u^2+q^2}du \\ &= \frac{A}{2}\ln(u^2+q^2) + B\int \frac{1}{u^2+q^2}du.\end{aligned}$$

The last antiderivative can be found by a trigonometric substitution.

17.60 Example. To find $\displaystyle\int \frac{u}{4u^2+8u+7}du$:
Let

$$\begin{aligned}I = \int \frac{u}{4u^2+8u+7}du &= \frac{1}{4}\int \frac{u}{u^2+2u+\frac{7}{4}}du \\ &= \frac{1}{4}\int \frac{u}{u^2+2u+1+\frac{3}{4}}du \\ &= \frac{1}{4}\int \frac{u}{(u+1)^2+\frac{3}{4}}du.\end{aligned}$$

Let $t = u + 1$, so $u = t - 1$ and $du = dt$. Then

$$\begin{aligned}I &= \frac{1}{4}\int \frac{t-1}{t^2+\frac{3}{4}}dt = \frac{1}{8}\int \frac{2t}{t^2+\frac{3}{4}}dt - \frac{1}{4}\int \frac{1}{t^2+\frac{3}{4}}dt \\ &= \frac{1}{8}\ln(t^2+\frac{3}{4}) - \frac{1}{4}\int \frac{1}{t^2+\frac{3}{4}}dt \\ &= \frac{1}{8}\ln\left((u+1)^2+\frac{3}{4}\right) - \frac{1}{4}\int \frac{1}{t^2+\frac{3}{4}}dt \\ &= \frac{1}{8}\ln(u^2+2u+\frac{7}{4}) - \frac{1}{4}\int \frac{1}{t^2+\frac{3}{4}}dt.\end{aligned}$$

Now let $t = \frac{\sqrt{3}}{2}\tan\theta$, so $dt = \frac{\sqrt{3}}{2}\sec^2\theta\, d\theta$, and $t^2 + \frac{3}{4} = \frac{3}{4}\sec^2\theta$. Then

$$\int \frac{1}{t^2+\frac{3}{4}}dt = \int \frac{\frac{\sqrt{3}}{2}\sec^2\theta}{\frac{3}{4}\sec^2\theta}d\theta = \frac{2}{\sqrt{3}}\int d\theta$$

$$\begin{aligned} &= \frac{2}{\sqrt{3}}\theta = \frac{2}{\sqrt{3}}\arctan\Big(\frac{2t}{\sqrt{3}}\Big) \\ &= \frac{2}{\sqrt{3}}\arctan\Big(\frac{2u+2}{\sqrt{3}}\Big). \end{aligned}$$

Hence,

$$I = \frac{1}{8}\ln\Big(u^2+2u+\frac{7}{4}\Big) - \frac{1}{2\sqrt{3}}\arctan\Big(\frac{2u+2}{\sqrt{3}}\Big).$$

To find $\int \frac{R(x)}{x^2+ax+b}dx$ where R is a polynomial of degree > 1.
First use long division to write

$$\frac{R(x)}{x^2+ax+b} = Q(x) + \frac{P(x)}{x^2+ax+b}$$

where Q is a polynomial, and P is a polynomial of degree ≤ 1. Then use one of the methods already discussed.

17.61 Example. To find $\int \frac{x^3+1}{x^2+1}dx$. By using long division, we get

$$\begin{array}{r|rrr} & x & & \\ \hline x^2+1 & x^3 & & +1 \\ & x^3 & +x & \\ \hline & & -x & +1 \end{array}$$

$$\frac{x^3+1}{x^2+1} = x + \frac{-x+1}{x^2+1}.$$

Hence

$$\begin{aligned} \int \frac{x^3+1}{x^2+1}dx &= \int x - \frac{x}{x^2+1} + \frac{1}{x^2+1}dx \\ &= \frac{1}{2}x^2 - \frac{1}{2}\ln(x^2+1) + \arctan(x). \end{aligned}$$

17.62 Example. In exercise 17.7, you showed that $\ln\Big(|\sec(x)+\tan(x)|\Big)$ is an antiderivative for $\sec(x)$. The function $\ln\Big(|\sec(x)+\tan(x)|\Big)$ in that exercise appeared magically with no motivation. I will now derive the formula, using standard methods:

$$\int \sec(x)\,dx = \int \frac{1}{\cos(x)}\,dx = \int \frac{\cos(x)}{\cos^2(x)}\,dx = \int \frac{\cos(x)}{1-\sin^2(x)}\,dx.$$

Now let $u = \sin(x)$. Then $du = \cos(x)\,dx$, and

$$\int \sec(x)\,dx = \int \frac{du}{1-u^2}.$$

Suppose $\dfrac{1}{1-u^2} = \dfrac{A}{1-u} + \dfrac{B}{1+u}$. Then

$$\frac{1}{1+u} = A + \frac{B(1-u)}{(1+u)},$$

and if we take the limit of both sides as $u \to 1$ we get $A = \dfrac{1}{2}$. Also

$$\frac{1}{1-u} = \frac{A(1+u)}{1-u} + B,$$

and if we take the limit as $u \to -1$, we get $B = \dfrac{1}{2}$. Thus

$$\begin{aligned}
\int \sec(x)\,dx &= \int \frac{1}{1-u^2}\,du \\
&= \frac{1}{2}\int \left(\frac{1}{1-u} + \frac{1}{1+u}\right)\,du \\
&= \frac{1}{2}\left[-\ln(|1-u|) + \ln(|1+u|)\right] \\
&= \frac{1}{2}\ln\left(\left|\frac{1+u}{1-u}\right|\right) \\
&= \frac{1}{2}\ln\left(\left|\frac{1+\sin(x)}{1-\sin(x)}\right|\right).
\end{aligned}$$

Now

$$\frac{1+\sin(x)}{1-\sin(x)} = \frac{1+\sin(x)}{1-\sin(x)}\cdot\frac{1+\sin(x)}{1+\sin(x)} = \frac{\Big(1+\sin(x)\Big)^2}{1-\sin^2(x)} = \frac{\Big(1+\sin(x)\Big)^2}{\cos^2(x)},$$

so

$$\begin{aligned}\frac{1}{2}\ln\left(\left|\frac{1+\sin(x)}{1-\sin(x)}\right|\right) &= \frac{1}{2}\ln\left(\left|\frac{1+\sin(x)}{\cos(x)}\right|^2\right) = \ln\left(\left|\frac{1+\sin(x)}{\cos(x)}\right|\right)\\ &= \ln\Big(|\sec(x)+\tan(x)|\Big),\end{aligned}$$

and thus

$$\int \sec(x)\,dx = \ln\Big(|\sec(x)+\tan(x)|\Big).$$

17.63 Exercise. Criticize the following argument:

I want to find $\displaystyle\int \frac{x^2}{x^2-1}dx = \int \frac{x^2}{(x-1)(x+1)}dx$. Suppose

$$\frac{x^2}{(x-1)(x+1)} = \frac{A}{x-1} + \frac{B}{x+1}.$$

Then

$$\frac{x^2}{x+1} = A + \frac{(x-1)B}{x+1}.$$

If we take the limit of both sides as $x \to 1$, we get $\dfrac{1}{2} = A$. Also

$$\frac{x^2}{x-1} = \frac{A(x+1)}{x-1} + B,$$

and if we take the limit of both sides as $x \to -1$, we get $-\dfrac{1}{2} = B$. Thus

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

Hence,

$$\int \frac{x^2}{x^2-1}dx = \frac{1}{2}\ln(|x-1|) - \frac{1}{2}\ln(|x+1|).$$

17.64 Exercise. Find the following antiderivatives:

a) $\displaystyle\int \frac{1}{4x^2-1}dx$

b) $\displaystyle\int \frac{1}{4x^2+1}dx$

c) $\displaystyle\int \frac{x+1}{x^2-6x+8}dx$

d) $\displaystyle\int \frac{x+1}{x^2-6x+9}dx$

e) $\displaystyle\int \frac{1}{9x^2+6x+2}dx$

f) $\displaystyle\int \frac{x^3}{x^2+1}dx$

g) $\displaystyle\int \frac{1}{\sqrt{x^2+2x+2}}dx$

17.65 Exercise. Find the following antiderivatives:

a) $\displaystyle\int \frac{\cos(ax)}{\sin^3(ax)}dx.$

b) $\displaystyle\int \frac{\sin(t)\cos(t)}{\cos^2(t)+1}dt.$

c) $\displaystyle\int \frac{1}{(1-t)^3}dt.$

d) $\displaystyle\int \frac{1}{5+4x+x^2}dx.$

e) $\displaystyle\int x^3\sqrt{x^2+1}dx.$

f) $\displaystyle\int \frac{1}{\sqrt{-3-4x-x^2}}.$

g) $\displaystyle\int \frac{\sin(2\theta)}{\cos^2(\theta)-\sin^2(\theta)}d\theta.$

h) $\displaystyle\int (1+\tan(u))^2du.$

i) Choose a number p, and find $\displaystyle\int x^p(x^{10}-2)^{10}dx.$

j) Choose a number q, and find $\displaystyle\int x^q e^{-\frac{1}{x}}dx.$

k) $\int xe^{-x^2}dx.$

l) $\int \frac{u^3}{1+u^2}du.$

m) $\int x^2 \arctan(x)dx.$

n) $\int x^3(1+x)^{\frac{1}{4}}dx.$

o) $\int xe^{2x}dx.$

p) $\int \arcsin(x)dx.$

Bibliography

[1] Tom M. Apostol. *Calculus*, volume I. John Wiley & Sons, Inc., second edition, 1967.

[2] Archimedes. *The Works of Archimedes Edited in Modern Notation with Introductory Chapters by T. L. Heath.* Dover Publications, 195? QA31 .A692.

[3] Archimedes. *Les Œuvres complètes d'archimède. Traduites du grek en français par Paul Ver Eecke.* Vaillant-Carmanne S. A., 1960. QA31 .A693 1960.

[4] Walter William Rouse Ball. *Mathematical Recreations and Essays, rev. by H. S. M. Coxeter.* Macmillan Co., 1947. QA95 .B17 1939.

[5] Stefan Banach. Sur le problème de la mesure. *Fundamenta Mathematicae*, 4:7–33, 1923.

[6] Margaret E. Baron. *The Origins of the Infinitesimal Calculus.* Dover Publications, 1987.

[7] George Berkeley. *The Works of George Berkeley Bishop of Cloyne Edited by A. A. Luce and T. E. Jessop.* Thomas Nelson and Sons, 1951. B1304.L8 1948 v.4 *The Analyst* appears in volume 4.

[8] Jacob Bernoulli. *Die Werke Von Jacob Bernoulli*, volume 3. Birkäuser Verlag Basel, 1975. QA3 B52 1969 v.3.

[9] Jakob Bernoulli. Sequences and series. In D. J. Struik, editor, *A Source Book in Mathematics 1200–1800*, pages 316–320. Harvard University Press, 1969. QA21 .S88.

[10] Errett Bishop. Schizophrenia in contemporary mathematics. In Murray Rosenblatt, editor, *Errett Bishop: Reflections on Him and His Research*, pages 1–32. American Mathematical Society, 1985. QA326 .E75 1985.

[11] Salomon Bochner. *The Role of Mathematics in the Rise of Science.* Princeton University Press, 1966. QA21 .B6.

[12] J. M. Borwein, P. B. Borwein, and D. H. Bailey. Ramanujan, modular equations, and approximations to pi or how to compute one billion digits of pi. *American Mathematical Monthly*, 96:201–219, 1989.

[13] Carl Benjamin Boyer. *A History of Mathematics.* Princeton University Press, 1985. QA21 .B767 1985.

[14] R. Creighton Buck. *Advanced Calculus.* McGraw Hill, third edition, 1978. QA303 .B917 1978.

[15] Florian Cajori. *A History of Mathematical Notations(Two volumes bound as one).* Dover Publications, 1993. QA41 .C32 1993

[16] Arnold Buffum Chace, editor. *The Rhind Mathematical Papyrus : Free Translation and Commentary by Arnold Buffum Chace.* Mathematical Association of America, 1927–1929. Oversize QA30 .P21 v.1 Main Stacks.

[17] Euclid. *Euclid's Elements Translated with Introduction and Commentary by Sir Thomas L. Heath (3 volumes).* Dover Publications, 1956. QA31 .E875 1956.

[18] Leonard Euler. *Introduction to the Analysis of the Infinite Book I Translated by John D. Blanton.* Springer–Verlag, 1988. QA295 .E8413 1988.

[19] Pierre de Fermat. Applications of the geometric progression to the quadrature of parabolas and infinite hyperbolas. In D. J. Struik, editor, *A Source Book in Mathematics 1200–1800*, pages 219–222. Harvard University Press, 1969. QA21 .S88.

[20] Robert M. French. The Banach Tarski theorem. *Mathematical Intelligencer*, 10:21–28, 1988.

[21] Galileo Galilei. *Dialogues Concerning Two New Sciences, Translated by Henry Crew & Alfonso de Salvio.* Dover Publications, 1954. Library has edition published by Macmillan, QC17 .G15.

[22] Herman H. Goldstine. *A History of Numerical Analysis from the 16th through the 19th Century*. Springer-Verlag, 1977. QA297 .G64.

[23] Judith V. Grabiner. *The Origins of Cauchy's Rigorous Calculus*. MIT Press, 1981.

[24] Felix Hausdorff. *Grundzüge der Mengenlehre*. Chelsea Publishing Company, 1949. QA248 .H36G.

[25] Sir Thomas Heath. *Mathematics in Aristotle*. Oxford University Press, 1949. QA31 A95H4.

[26] Sir Thomas Heath. *A History of Greek Mathematics*, volume 2. Dover Publications, 1981. QA22 .H4 1981 v.2.

[27] George Gheverghese Joseph. *The Crest of the Peacock : Non-European Roots of Mathematics*. I.B. Tauris, 1991. QA21 .G53 1991.

[28] H. Jerome Keisler. The hyperreal line. In Philip Ehrlich, editor, *Real Numbers, Generalizations of the Reals, and Theories of Continua*. Kluwer Academic Publishers, 1994. QA241 .R34 1994.

[29] Morris Kline. *Mathematical Thought from Ancient to Modern Times*. Oxford University Press, 1972. QA21 .53.

[30] Donald Ervin Knuth. *Surreal numbers : How Two Ex-students Turned on to Pure Mathematics and Found Total Happiness : a mathematical novelette*. Addison-Wesley Pub. Co., 1974. QA241 .K688.

[31] Helge von Koch. Une méthod géometrique Élémentaire pour l'Étude de certaines questiones de la théorie des courbes planes. *Acta Mathematica*, 30:145–174, 1906.

[32] Joseph D. E. Konhauser. $\int \sec^3 x dx$. *Mathematics Magazine*, 38:45, 1965.

[33] Gottfried Leibniz. A new method for maxima and minima as well as tangents, which is neither impeded by fractional nor irrational quantities, and a remarkable type of calculus for them. In D. J. Struik, editor, *A Source Book in Mathematics 1200–1800*, pages 272–280. Harvard University Press, 1969. QA21 .S88.

[34] Gottfried Wilhelm Leibniz. *The Early Mathematical Manuscripts of Leibniz; tr. from the Latin texts published by Carl Immanuel Gerhardt with critical and historical notes, by J. M. Child.* The Open Court Publishing Company, 1920. QA3 .L52E.

[35] Guillaume François de l'Hôpital. The analysis of the infinitesimally small. In D. J. Struik, editor, *A Source Book in Mathematics 1200–1800*, pages 312–316. Harvard University Press, 1969. QA21 .S88.

[36] Kenneth S. Miller and Bertram Ross. *An Introduction to the Fractional Calculus and Fractional Differential Equations.* John Wiley & Sons, Inc, 1993. QA 303 .M6813 1993.

[37] Roger B. Nelsen. *Proofs Without Words.* Mathematical Association of America, 1993. QA9.54 .N4 1993.

[38] Otto Neugebauer. *The Exact Sciences in Antiquity.* Dover Publications, 1969. The Reed math library has an edition published by Brown University Press, QA22 .N36 1957.

[39] Bernhard Riemann. *Gesammelte Mathematische Werke.* B. G. Teubner, second edition, 1892. QA3.R55 W.

[40] Joseph Fels Ritt. *Integration in Finite Terms; Liouville's Theory of Elementary Methods.* Columbia University Press, 1948. QA308 .R61.

[41] Maxwell Rosenlicht. Integration in finite terms. *American Mathematical Monthly*, 79:963–972, 1972.

[42] David Eugene Smith. *History of Mathematics*, volume II. Dover Publications, 1958. QA21 .S62 1958 v.2.

[43] David Eugene Smith, editor. *A Source Book in Mathematics.* Dover Publications, 1959. QA3 .S66 1959.

[44] Michael Spivak. *Calculus.* Publish or Perish, Inc., second edition, 1980. Library has first edition QA303 .S78.

[45] B. L. van der Waerden. *Science Awakening, English translation of* Ontwakende Wetenschap *by Arnold Dresden, with additions by the author.* Oxford University Press, 1961. QA22 .W314 1961.

[46] François Marie Arouet de Voltaire. Philosophical dictionary. In *The works of Voltaire : a contemporary version / a critique & biography by John Morley ; notes by Tobias Smollett ; rev. and modernized new translations by William F. Fleming and an introduction by Oliver H.G. Leigh*, volume 5,part II. E. R. DuMont, 1901. PQ2070 1901 v.5.

[47] Lǐ Yan and Dù Shíràn. *Chinese Mathematics A Concise History translated by John N. Crossley and Anthony W. -C. Lun.* Oxford University Press, 1987.

Appendix A

Hints and Answers

Exercise 3: The Rhind value is $256/81 = 3.1604\ldots$

Exercise 1.7: Look at the boundary.

Exercise 1.10: If a set has no endpoints, then it contains all of its endpoints and none of its endpoints.

Exercise 2.10: $1^3 + 2^3 + \cdots n^3 = \frac{n^2(n+1)^2}{4}$.

Exercise 2.18: $\text{area}T(a) = \dfrac{2}{3}a^{\frac{3}{2}}$.

Exercise 2.27: $.027027027\ldots = \frac{1}{37}$.

Exercise 2.36: I let $O_j = B(a^{\frac{j}{N}}, a^{\frac{j-1}{N}}; 0, a^{-\frac{2j}{N}})$ and $I_j = B(a^{\frac{j}{N}}, a^{\frac{j-1}{N}}; 0, a^{-\frac{2(j-1)}{N}})$.

Exercise 3.20: Recall $(R \Longrightarrow S) \Longleftrightarrow (S$ or not $R)$.

Exercise 5.61: $S_1^{ab}[\frac{1}{t}] = S_1^a[\frac{1}{t}] \cup S_a^{ab}[\frac{1}{t}]$. (Draw a picture.)

Exercise 5.80: Consider a partition with a fairly large number of points.

Exercise 6.33: The assertion is false.

Exercise 6.59: (part e) The limit is $\frac{1}{3}$. It simplifies matters if you factor both the numerator and the denominator. The sequence in part g) is a translate of the sequence in part f).

Exercise 6.69: All four statements are false.

Exercise 6.94: a) $(1+\frac{3}{n})^{2n} = ((1+3/n)^n)^2$.

Exercise 6.97: $(1-\frac{c}{n})^n = \frac{(1-\frac{c^2}{n^2})^n}{(1+\frac{c}{n})^n}$.

Exercise 7.16: Take $c = \frac{b}{a}$ in lemma 7.13.

Exercise 7.18: $A_0^a f = A_0^{\frac{1}{n}} f + A_{\frac{1}{n}}^a f$. Show that $A_0^{\frac{1}{n}} f$ is small when n is large.

Exercise 8.14: e)$\frac{x+1}{x} = 1 + \frac{1}{x}$. Not all of these integrals exist.

Exercise 8.16: Show that $\sum(f,P,S) \leq \sum(g,P,S)$ for every partition P of $[a,b]$ and every sample S for P.

Exercise 8.28: g is the sum of an integrable function and a spike function.

Exercise 8.32: f is not piecewise monotonic. It is easy to see that f is integrable on $[1,2]$. If you can show it is integrable on $[0,1]$then you are essentially done.

Exercise 8.34: b) $(b-a)^3/6$.

Exercise 8.41: For any partition P of $[0.1]$ you can find a sample S such that $\sum(R,P,S) = 0$.

Exercise 8.46: In equation 8.44, replace r by $\frac{1}{R}$, , and replace a and b by RA and RB.

Exercise 8.48: $\alpha(E_{ab}) = \pi ab$.

Exercise 8.50: If $a = 1/4$ then both areas are approximately 3.1416

Exercise 8.55: area $= 4\pi$.

Exercise 8.57: The areas are $5/12$ and 1.

Exercise 8.58: The area is $\frac{37}{12}$. Some fractions with large numerators may appear along the way.

Exercise 9.20: The last two formulas are obtained from the second by replacing t by $t/2$.

Exercise 9.29: I used exercise 9.28 with $x = \frac{\pi}{6}$ to find $\cos(\frac{\pi}{6})$. You can also give a more geometric proof.

Exercise 9.44: You will need to use (9.24).

Exercise 9.48: $\alpha(S_0^\pi(\sin)) = 2$.

Exercise 9.49: area $= \sqrt{2}$.

Exercise 9.69: $\int_0^{\pi/2} \sin(x)dx = 1$; $\int_0^{\pi/2} \sin^2(x)dx = \pi/4$; $\int_0^{\pi/2} \sin^4(x)dx = 3\pi/16$.

Exercise 10.25: $f'(a) = -\frac{1}{a^2}$.

Exercise 10.26: See example 10.9 and 9.26.

Exercise 10.27: $f'(a) = \frac{1}{(a+1)^2}$.

Exercise 10.28: $y = 2x - 4$; $y = -6x - 4$.

Exercise 11.6: I used formula 9.25

Exercise 11.15: $\frac{d}{dt}(| - 100t|) = \frac{100t}{|t|}$.

Exercise 11.21: You can use the definition of derivative, or you can use the product rule and the reciprocal rule.

Exercise 11.24: $f'(x) = \ln(x)$, $g'(x) = \frac{ad-bc}{(cx+d)^2}$, $k'(x) = 2(2x+3)(x^2+3x+11)$

Exercise 11.29: $(g \circ (g \circ g))(x) = ((g \circ g) \circ g)(x) = x$ for $x \in \mathbf{R} \setminus 0, 1$. If you said $(f \circ f)(x) = x$, calculate both sides when $x = -1$.

Exercise 11.40: Use the definition of derivative. $h'(2) = 0$.

Exercise 11.43: $g'(x) = -\tan(x)$, $h'(x) = \tan(x)$, $k'(x) = \sec(x)$, $l'(x) = -\csc(x)$, $m'(x) = 9x^2 \ln(5x)$, $n'(x) = \frac{\sqrt{x^2+1}}{x}$ (It requires a lot of calculation to simplify n'), $p'(x) = \frac{x^2}{x+4}$, $q'(x) = \sin(\ln(|6x|))$.

Exercise 12.14: d) Such a function k does exist.

Exercise 12.15: a) Use extreme value property.

Exercise 12.27: Proof is like given proof of corollary 12.26.

Exercise 12.31: Apply corollary 12.26 to $F - G$.

Exercise 12.35: Yes.

Exercise 12.36: You can apply the chain rule to the identity $f(-x) = f(x)$.

Exercise 13.14: The function to minimize is $f(x) =$distance$((0, \frac{9}{2}), (x, x^2))$.

Exercise 13.15: You may get a complicated equation of the form $\sqrt{f(x)} = \sqrt{g(x)}$ to solve. Square both sides and the equation should simplify.

Exercise 14.5: Apply the intermediate value property to $f - fp$.

Exercise 14.9: One of the zeros is in $[1, 2]$.

Exercise 14.10: I showed that if temp$(A) <$ temp$(B) <$ temp(D), then there is a point Q in $DC \cup CA$ such that temp$(Q) =$ temp(B).

Exercise 14.11: if temp$(A) <$ temp$(B) <$ temp$(C) <$ temp(D), find two points different from B that have the same temperature as B.

Exercise 14.17: You may want to define some of these functions using more than one formula.

Exercise 14.41: Use the extreme value property to get A and B.

Exercise 14.54: $f'(x) = 2\sqrt{a^2 - x^2}$; $h'(x) = \arccos(ax)$; $n'(x) = (a^2+b^2)e^{ax}\sin(bx)$; $p'(x) = a^3x^2e^{ax}$.

Exercise 14.55: It is not true that $l(x) = x$ for all x. Note that the image of l is $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Exercise 15.5: $g^{(k)}(t) = tf^{(k)}(t) + kf^{(k-1)}(t)$.

Exercise 15.8: Use the antiderivative theorem twice.

Exercise 15.9: $(fg)^{(3)} = fg^{(3)} + 3f^{(1)}g^{(2)} + 3f^{(2)}g^{(1)} + f^{(3)}g$.

Exercise 15.13: You will need to use a few trigonometric identities, including the reflection law (9.18).

Exercise 15.22: $h(t) = h_0 + v_0t - \frac{1}{2}gt^2$.

Exercise 15.29: Use theorem 15.27 and corollary 12.26

Exercise 16.2: You probably will not be able to find a "single formula" for this. My function has a local maximum at $\frac{1}{2n+1}$ for all $n \in \mathbf{Z}^+$.

Exercise 16.8: The result is known if $p < q$. To get the result when $q < p$, apply 16.6 to f on $[q, p]$.

Exercise 16.13: Not all of these integrals make sense. $K'(x) = 1$ for all $x \in \mathbf{R}$. $L'(x) = 1$ for $x \in \mathbf{R}^+$. $L'(x) = -1$ for $x \in \mathbf{R}^-$. $L(0)$ is not defined.

Exercise 17.16: b) $-\ln(|\csc(e^x) + \cot(e^x)|)$; f) $\frac{1}{2}\ln(|\sin(2x)|)$. i) Cf example 9.68i.j) You did this in exercise 9.69.

Exercise 17.31: b) $\frac{1}{2}e^x(\sin(x) - \cos(x))$. When you do the second integration by parts, be careful not to undo the first. c) $x\arctan(x) - \frac{1}{2}\ln(1 + x^2)$. Let $g'(x) = 1$. d) and e) can be done easily without using integration by parts. f) If $r = -1$ the answer is $\frac{1}{2}(\ln(|x|))^2$.

Exercise 17.42: c) Let $u = \sqrt{x}$. You will need an integration by parts. d) Let $u = \ln(3x)$. e) Remember the definition of 2^x.

Exercise 17.49: a) $\frac{1}{2}a^2\arcsin(\frac{x}{a}) + \frac{1}{2}x\sqrt{a^2 - x^2}$. If you forget how to find $\int \cos^2(\theta)d\theta$, review example 9.53. Also recall that $\sin(2x) = 2\sin(x)\cos(x)$. b) $\ln(\frac{x+\sqrt{a^2+x^2}}{a})$. c) and d) do not require a trigonometric substitution.

Exercise 17.53: a) $\frac{5}{4}$. b) $\frac{\pi}{6}$. c)$\frac{4}{15}$.

Exercise 17.54: $\frac{2\pi}{3} + \sqrt{3}$.

Exercise 17.64: (g) $\ln(\sqrt{x^2 + 2x + 2} + x + 1)$. First complete the square, and then reduce the problem to $\int \frac{1}{\sqrt{u^2+1}}du$.

Appendix B

Proofs of Some Area Theorems

B.1 Theorem (Addition Theorem.) *For any bounded sets S and T in $\mathbf{R}^2$*

$$\alpha(S \cup T) = \alpha(S) + \alpha(T) - \alpha(S \cap T). \tag{B.2}$$

and consequently

$$\alpha(S \cup T) \leq \alpha(S) + \alpha(T).$$

Proof: We have

$$S \cup T = S \cup (T \setminus S) \text{ where } S \cap (T \setminus S) = \emptyset$$

and

$$T = (T \setminus S) \cup (T \cap S) \text{ where } (T \setminus S) \cap (T \cap S) = \emptyset.$$

Hence by the additivity of area

$$\alpha(S \cup T) = \alpha(S) + \alpha(T \setminus S) \tag{B.3}$$

and

$$\alpha(T) = \alpha(T \setminus S) + \alpha(T \cap S) \tag{B.4}$$

If we solve equation (B.4) for $\alpha(T \setminus S)$ and use this result in equation (B.3) we get the desired result. ‖

B.5 Corollary (Subadditivity of area.) *Let $n \in \mathbf{Z}_{\geq 1}$, and let A_1, A_2, $\cdots$ A_n be bounded sets in $\mathbf{R}^2$. Then*

$$\alpha(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \alpha(A_i). \tag{B.6}$$

Proof: The proof is by induction. If $n = 1$, then (B.6) says $\alpha(A_1) \leq \alpha(A_1)$, which is true. Suppose now that k is a generic element of $\mathbf{Z}_{\geq 1}$, and that (B.6) is true when $n = k$. Let $A_1, \cdots, A_{k+1}$ be bounded sets in $\mathbf{R}^2$. Then

$$\begin{aligned}
\alpha(\bigcup_{i=1}^{k+1} A_i) &= \alpha\left(\bigcup_{i=1}^{k} A_i \cup A_{k+1}\right) \\
&\leq \alpha\left(\bigcup_{i=1}^{k} A_i\right) + \alpha(A_{k+1}) \\
&\leq \sum_{i=1}^{k} \alpha(A_i) + \alpha(A_{k+1}) = \sum_{i=1}^{k+1} \alpha(A_i).
\end{aligned}$$

Hence (B.6) is true when $n = k + 1$, and by induction the formula holds for all $n \in \mathbf{Z}_{n \geq k}$. ‖

B.7 Theorem (Monotonicity of Area.) *Let S, T be bounded sets such that $S \subset T$. Then $\alpha(S) \leq \alpha(T)$.*

Proof: If $S \subset T$ then $S \cap T = S$, and in this case equation (B.4) becomes

$$\alpha(T) = \alpha(T \setminus S) + \alpha(S).$$

Since $\alpha(T \setminus S) \geq 0$, it follows that $\alpha(T) \geq \alpha(S)$. ‖

B.8 Theorem (Additivity for almost disjoint sets.) *Let $\{R_1, \cdots, R_n\}$ be a finite set of bounded sets such that R_i and R_j are almost disjoint whenever $i \neq j$. Then*

$$\alpha(\bigcup_{i=1}^{n} R_i) = \sum_{i=1}^{n} \alpha(R_i). \tag{B.9}$$

Proof: The proof is by induction on n. For $n = 1$, equation (B.9) says that $\alpha(R_1) = \alpha(R_1)$, and this is true. Now suppose $\{R_1 \cdots R_{n+1}\}$ is a family of mutually almost-disjoint sets. Then

$$(R_1 \cup \cdots \cup R_n) \cap R_{n+1} = (R_1 \cap R_{n+1}) \cup (R_2 \cap R_{n+1}) \cup \cdots \cup (R_n \cap R_{n+1})$$

and this is a finite union of zero-area sets, and hence is a zero-area set. Hence, by the addition theorem,

$$\alpha((R_1 \cup \cdots \cup R_n) \cup R_{n+1}) = \alpha(R_1 \cup \cdots \cup R_n) + \alpha(R_{n+1})$$

i.e.,

$$\alpha(\bigcup_{i=1}^{n+1} R_i) = \sum_{i=1}^{n} \alpha(R_i) + \alpha(R_{n+1}) = \sum_{i=1}^{n+1} \alpha(R_i).$$

The theorem now follows from the induction principle. ‖

Appendix C

Prerequisites

C.1 Properties of Real Numbers

Algebraic Laws

Commutative laws for addition and multiplication: If a and b are arbitrary real numbers then

$$a + b = b + a, \tag{C.1}$$
$$ab = ba. \tag{C.2}$$

Associative laws for addition and multiplication: If a, b, and c are arbitrary real numbers then

$$(a + b) + c = a + (b + c), \tag{C.3}$$
$$(ab)c = a(bc). \tag{C.4}$$

As a consequence of equations C.3 and C.4 we usually omit the parentheses in triple sums or products, and write $a + b + c$ or abc. We know that all meaningful ways of inserting parentheses yield the same result.

Distributive laws: If a, b and c are arbitrary real numbers, and d is an arbitrary non-zero real number then

$$c(a + b) = ca + cb, \tag{C.5}$$
$$c(a - b) = ca - cb, \tag{C.6}$$
$$(a + b)c = ac + bc, \tag{C.7}$$

$$(a-b)c = ac - bc, \tag{C.8}$$
$$(a+b)/d = a/d + b/d, \tag{C.9}$$
$$(a-b)/d = a/d - b/d. \tag{C.10}$$

Properties of zero and one: The rational numbers 0 and 1, have the property that for all real numbers a

$$a + 0 = a, \tag{C.11}$$
$$0 + a = a, \tag{C.12}$$
$$a \cdot 1 = a, \tag{C.13}$$
$$1 \cdot a = a, \tag{C.14}$$
$$0 \cdot a = 0, \tag{C.15}$$
$$a \cdot 0 = 0. \tag{C.16}$$

Moreover

$$0 \neq 1, \tag{C.17}$$

and

$$\text{if } \; ab = 0 \text{ then } \; a = 0 \text{ or } \; b = 0 \text{ (or both).} \tag{C.18}$$

Additive and multiplicative inverses: For each real number a there is a real number $-a$ (called the *additive inverse of* a) and for each non-zero real number b there is a real number b^{-1} (called *the multiplicative inverse of* b) such that

$$a + (-a) = 0, \tag{C.19}$$
$$(-a) + a = 0, \tag{C.20}$$
$$b \cdot b^{-1} = 1, \tag{C.21}$$
$$b^{-1} \cdot b = 1, \tag{C.22}$$
$$-0 = 0, \tag{C.23}$$
$$1^{-1} = 1. \tag{C.24}$$

Moreover for all real numbers a, c and all non-zero real numbers b

$$-(-a) = a, \tag{C.25}$$
$$a - c = a + (-c), \tag{C.26}$$
$$a/b = a \cdot b^{-1}, \tag{C.27}$$

$$b^{-1} = 1/b, \tag{C.28}$$
$$(ab)^{-1} = a^{-1}b^{-1} \tag{C.29}$$
$$-a = (-1)\cdot a, \tag{C.30}$$
$$(b^{-1})^{-1} = b, \tag{C.31}$$
$$(-a)(-c) = ac, \tag{C.32}$$
$$(-a)c = a(-c) = -(ac), \tag{C.33}$$
$$-(\frac{a}{b}) = \frac{-a}{b} = \frac{a}{-b}. \tag{C.34}$$

Note that by equation C.33, the expression $-xy$ without parentheses is unambiguous, i.e. no matter how parentheses are put in the result remains the same.
Cancellation laws: Let a, b, c be real numbers. Then

$$\text{if } a+b=a+c, \quad \text{then} \quad b=c. \tag{C.35}$$
$$\text{if } b+a=c+a, \quad \text{then} \quad b=c. \tag{C.36}$$
$$\text{if } ab=ac \text{ and } a\neq 0 \quad \text{then} \quad b=c. \tag{C.37}$$
$$\text{if } ba=ca \text{ and } a\neq 0 \quad \text{then} \quad b=c. \tag{C.38}$$

Some miscellaneous identities: For all real numbers a, b, c, d, x

$$a^2-b^2 = (a-b)(a+b), \tag{C.39}$$
$$(a+b)^2 = a^2+2ab+b^2, \tag{C.40}$$
$$(a-b)^2 = a^2-2ab+b^2, \tag{C.41}$$
$$(x+a)(x+b) = x^2+(a+b)x+ab, \tag{C.42}$$
$$(a+b)(c+d) = ac+ad+bc+db. \tag{C.43}$$

Moreover, if $b \neq 0$ and $d \neq 0$ then

$$\frac{a}{b}\cdot\frac{c}{d} = \frac{ac}{bd}. \tag{C.44}$$
$$\frac{a}{b}+\frac{c}{d} = \frac{ad+bc}{bd}. \tag{C.45}$$
$$\frac{a}{b}-\frac{c}{d} = \frac{ad-bc}{bd}. \tag{C.46}$$

If w, x, y, z are real numbers, then

$$w-x-y+z+y \text{ means } w+((-x)+((-y)+(z+y))) \tag{C.47}$$

i.e. the terms of the sum are associated from right to left. It is in fact true that all meaningful ways of introducing parentheses into a long sum yield the same result, and we will assume this. I will often make statements like

$$w - x - y + z + y = w + z - x \tag{C.48}$$

without explanation. Equation C.48 can be proved from our assumptions, as is shown below, but we will usually take such results for granted.

Proof of equation C.48. Let w, x, y, z be real numbers. Then

$$\begin{aligned} w - x - y + z + y &= w + ((-x) + ((-y) + (z + y))) \text{ by C.47} \\ &= w + ((-x) + ((-y) + (y + z))) \text{ by C.1} \\ &= w + ((-x) + (((-y) + y) + z)) \text{ by C.3} \\ &= w + ((-x) + (0 + z)) \text{ by C.20} \\ &= w + ((-x) + z) \text{ by C.12} \\ &= w + (z + (-x)) \text{ by C.1} \\ &= w + z - x \text{ by C.47.} \end{aligned}$$

Order Laws

There is a relation $<$ (*less than*) defined on the real numbers such that for each pair a, b of real numbers, the statement "$a < b$" is either true or false, and such that the following conditions are satisfied:

Trichotomy law: For each pair a, b of real numbers exactly one of the following statements is true:

$$a < b, \qquad a = b, \qquad b < a. \tag{C.49}$$

We say that a real number p is *positive* if and only if $p > 0$, and we say that a real number n is *negative* if and only if $n < 0$. Thus as a special case of the trichotomy law we have:

If a is a real number, then exactly one of the following statements is true:

$$a \text{ is positive}, \qquad a = 0, \qquad a \text{ is negative}. \tag{C.50}$$

Sign laws: Let a, b be real numbers. Then

$$\text{if } a > 0 \text{ and } b > 0 \quad \text{then} \quad ab > 0 \text{ and } a/b > 0, \tag{C.51}$$

$$\text{if } a < 0 \text{ and } b > 0 \quad \text{then} \quad ab < 0 \text{ and } a/b < 0, \tag{C.52}$$
$$\text{if } a > 0 \text{ and } b < 0 \quad \text{then} \quad ab < 0 \text{ and } a/b < 0, \tag{C.53}$$
$$\text{if } a < 0 \text{ and } b < 0 \quad \text{then} \quad ab > 0 \text{ and } a/b > 0, \tag{C.54}$$
$$\text{if } a > 0 \text{ and } b > 0 \quad \text{then} \quad a + b > 0, \tag{C.55}$$
$$\text{if } a < 0 \text{ and } b < 0 \quad \text{then} \quad a + b < 0. \tag{C.56}$$

Also,

$$a \text{ is positive if and only if } -a \text{ is negative,} \tag{C.57}$$

and

$$a \text{ is positive if and only if } a^{-1} \text{ is positive.} \tag{C.58}$$

$$\text{if } ab > 0 \text{ then either } (a > 0 \text{ and } b > 0) \text{ or } (a < 0 \text{ and } b < 0). \tag{C.59}$$

$$\text{if } ab < 0 \text{ then either } (a > 0 \text{ and } b < 0) \text{ or } (a < 0 \text{ and } b > 0). \tag{C.60}$$

$$\text{if } a/b > 0 \text{ then either } (a > 0 \text{ and } b > 0) \text{ or } (a < 0 \text{ and } b < 0). \tag{C.61}$$

$$\text{if } a/b < 0 \text{ then either } (a > 0 \text{ and } b < 0) \text{ or } (a < 0 \text{ and } b > 0). \tag{C.62}$$

It follows immediately from the sign laws that for all real numbers a

$$a^2 \geq 0 \text{ and } \quad \text{if } a \neq 0 \text{ then } a^2 > 0. \tag{C.63}$$

Here, as usual a^2 means $a \cdot a$.

Transitivity of <: Let a, b, c be real numbers. Then

$$\text{if } a < b \text{ and } b < c \text{ then } a < c. \tag{C.64}$$

We write $a \leq b$ as an abbreviation for "either $a < b$ or $a = b$", and we write $b > a$ to mean $a < b$. We also nest inequalities in the following way:

$$a < b \leq c = d < e$$

means

$$a < b \text{ and } b \leq c \text{ and } c = d \text{ and } d < e.$$

Addition of Inequalities: Let a, b, c, d be real numbers. Then

$$\text{if } a < b \text{ and } c < d \text{ then } a + c < b + d, \tag{C.65}$$
$$\text{if } a \le b \text{ and } c \le d \text{ then } a + c \le b + d, \tag{C.66}$$
$$\text{if } a < b \text{ and } c \le d \text{ then } a + c < b + d, \tag{C.67}$$
$$\text{if } a < b \text{ then } a - c < b - c, \tag{C.68}$$
$$\text{if } c < d \text{ then } -c > -d, \tag{C.69}$$
$$\text{if } c < d \text{ then } a - c > a - d. \tag{C.70}$$

Multiplication of Inequalities: Let a, b, c, d be real numbers.

$$\text{if } a < b \text{ and } c > 0 \text{ then } ac < bc, \tag{C.71}$$
$$\text{if } a < b \text{ and } c > 0 \text{ then } a/c < b/c, \tag{C.72}$$
$$\text{if } 0 < a < b \text{ and } 0 < c < d \text{ then } 0 < ac < bd, \tag{C.73}$$
$$\text{if } a < b \text{ and } c < 0 \text{ then } ac > bc, \tag{C.74}$$
$$\text{if } a < b \text{ and } c < 0 \text{ then } a/c > b/c, \tag{C.75}$$
$$\text{if } 0 < a \text{ and } a < b \text{ then } a^{-1} > b^{-1}. \tag{C.76}$$

Discreteness of Integers: If n is an integer, then there are no integers between n and $n + 1$, i.e. there are no integers k satisfying $n < k < n + 1$. A consequence of this is that

$$\text{If } k, n \text{ are integers, and } k < n + 1, \text{ then } k \le n. \tag{C.77}$$

If x and y are real numbers such that $y - x > 1$ then there is an integer n such that

$$x < n < y. \tag{C.78}$$

Archimedean Property: Let x be an arbitrary real number. Then

$$\text{there exists an integer } n \text{ such that } n > x. \tag{C.79}$$

Miscellaneous Properties

Names for Rational Numbers: Every rational number r can be written as a quotient of integers:

$$r = \frac{m}{n} \text{ where } m, n \text{ are integers and } n \ne 0,$$

and without loss of generality we may take $n > 0$. In general, a rational number has many different names, e.g. $\frac{2}{3}$, $\frac{-10}{-15}$, and $\frac{34}{51}$ are different names for the same rational number. If I say "let $x = \frac{2}{3}$", I mean let x denote the rational number which has "$\frac{2}{3}$" as one of its names. You should think of each rational number as a specific point on the line of real numbers. Let m, n, p, q be integers with $n \neq 0$ and $q \neq 0$. Then

$$\frac{m}{n} = \frac{p}{q} \text{ if and only if } mq = np. \tag{C.80}$$

If n and q are *positive,* then

$$\frac{m}{n} < \frac{p}{q} \text{ if and only if } mq < np. \tag{C.81}$$

Equations C.80 and C.81 hold for arbitrary real numbers m, n, p, q. It will be assumed that if you are given two rational numbers, you can decide whether or not the first is less that the second. You also know that the sum, difference, and product of two integers is an integer, and the additive inverse of an integer is an integer.

Absolute value: If x is a real number, then the *absolute value* of x, denoted by $|x|$, is defined by

$$|x| = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x < 0. \end{cases} \tag{C.82}$$

For all real numbers x and all positive numbers a we have

$$(|x| < a) \text{ if and only if } (-a < x < a), \tag{C.83}$$
$$(|x| \leq a) \text{ if and only if } (-a \leq x \leq a). \tag{C.84}$$

For all real numbers x, y, z with $z \neq 0$,

$$|x| = |-x| \tag{C.85}$$
$$-|x| \leq x \leq |x| \tag{C.86}$$
$$|xy| = |x| \cdot |y| \tag{C.87}$$
$$\left|\frac{x}{z}\right| = \frac{|x|}{|z|}. \tag{C.88}$$

Powers: If a is a real number, and n is a non-negative integer, then the *power* a^n is defined by the rules

$$a^0 = 1 \tag{C.89}$$
$$a^{n+1} = a^n \cdot a \text{ for } n \geq 0. \tag{C.90}$$

If a is a non-zero number and n is a negative integer, then a^n is defined by

$$a^n = (a^{-n})^{-1} = \frac{1}{a^{-n}}. \tag{C.91}$$

If a is a non-negative number and n is a positive integer, then $a^{\frac{1}{n}}$ is defined by

$$a^{\frac{1}{n}} \text{ is the unique non-negative number } b \text{ such that } b^n = a. \tag{C.92}$$

If a is a non-negative number and m is an arbitrary integer and n is a positive integer, then $a^{\frac{m}{n}}$ is defined by

$$a^{\frac{m}{n}} = \begin{cases} (a^{\frac{1}{n}})^m & \text{if } a > 0. \\ 0 & \text{if } a = 0 \text{ and } m > 0 \\ \text{undefined} & \text{if } a = 0 \text{ and } m < 0. \end{cases} \tag{C.93}$$

If m, n, p, q are integers such that $n \neq 0$ and $q \neq 0$ and $\frac{m}{n} = \frac{p}{q}$, then

$$(a^{\frac{1}{n}})^m = (a^m)^{\frac{1}{n}} = (a^p)^{\frac{1}{q}} = (a^{\frac{1}{q}})^p. \tag{C.94}$$

Monotonicity of Powers: If r is a positive rational number, and x and y are non-negative real numbers, then

$$x < y \text{ if and only if } x^r < y^r. \tag{C.95}$$

If r is a negative rational number, and x and y are positive real numbers, then

$$x < y \text{ if and only if } x^r > y^r. \tag{C.96}$$

If a is a positive real number greater than 1, and p and q are rational numbers, then

$$p < q \text{ if and only if } a^p < a^q. \tag{C.97}$$

If a is a positive real number less than 1, and p and q are rational numbers, then

$$p < q \text{ if and only if } a^p > a^q. \tag{C.98}$$

Laws of exponents: Let a and b be real numbers, and let r and s be rational numbers. Then the following relations hold whenever all of the powers involved are defined:

$$a^r a^s = a^{r+s}, \tag{C.99}$$
$$(a^r)^s = a^{(rs)}, \tag{C.100}$$
$$(ab)^r = a^r b^r. \tag{C.101}$$
$$a^{-r} = \frac{1}{a^r} \tag{C.102}$$

Remarks on equality: If x, y and z are names for mathematical objects, then we write $x = y$ to mean that x and y are different names for the same object. Thus

$$\text{if } \ x = y \text{ then } \ y = x, \tag{C.103}$$

and it is always the case that

$$x = x. \tag{C.104}$$

It also follows that

$$\text{if } \ x = y \text{ and } \ y = z \text{ then } \ x = z, \tag{C.105}$$

and more generally,

$$\text{if } \ x = y = z = t = w \text{ then } \ x = w. \tag{C.106}$$

If $x = y$, then the name x can be substituted for the name y in any statement containing the name x. For example, if x, y are numbers and we know that

$$x = y, \tag{C.107}$$

then we can conclude that

$$x + 1 = y + 1, \tag{C.108}$$

and that

$$x + x = x + y. \tag{C.109}$$

When giving a proof, one ordinarily goes from an equation such as C.107 to equations such as C.108 or C.109 without mentioning the reason, and the properties C.103–C.106 are usually used without mentioning them explicitly.

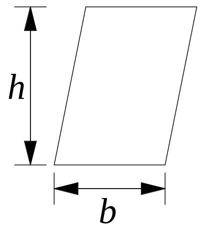
C.2 Geometrical Prerequisites

Area Formulas

It will be assumed that you are familiar with the results from Euclidean and coordinate geometry listed below.

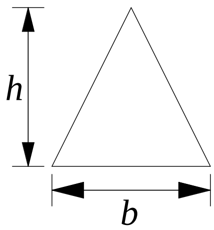
Area of a parallelogram:

$$\text{Area} = \text{base} \cdot \text{height} = bh \tag{C.110}$$



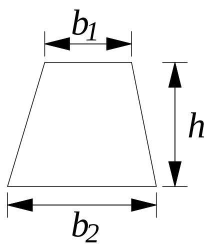
Area of a triangle:

$$\text{Area} = \tfrac{1}{2}\,\text{base} \cdot \text{height} = \tfrac{1}{2}bh \tag{C.111}$$



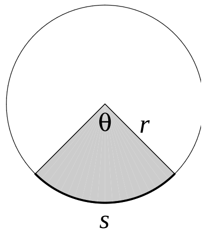
Area of a trapezoid:

$$\begin{aligned}\text{Area} &= (\text{average of bases}) \cdot \text{height} \\ &= \tfrac{1}{2}(b_1 + b_2)h\end{aligned} \tag{C.112}$$



We will always assume that angles are measured in radians unless otherwise specified. If an angle θ is inscribed in a circle of radius r and s is the length of the subtended arc, then

$$s = r\theta. \tag{C.113}$$



A right angle is $\pi/2$ and the sum of the angles of a triangle is π. When θ is four right angles in (C.113) we get

$$\text{circumference(circle)} = 2\pi r. \tag{C.114}$$

Area of a circular sector:

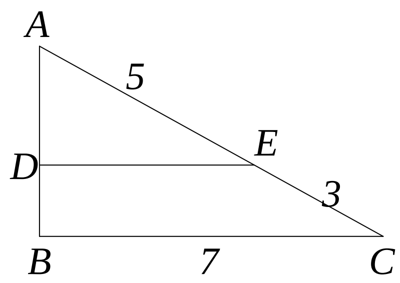
$$\begin{aligned} \text{Area} &= \frac{1}{2} \cdot \text{radius} \cdot \text{subtended arc} \\ &= \frac{1}{2} rs \\ &= \frac{1}{2} \cdot \text{central angle} \cdot \text{radius}^2 \\ &= \frac{1}{2} \theta r^2. \end{aligned} \tag{C.115}$$

In particular when θ is four right angles

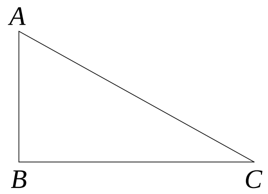
$$\text{Area(circle)} = \pi r^2. \tag{C.116}$$

Miscellaneous Properties

You should be familiar with the properties of parallel lines, and with the rules for deciding when triangles are congruent or similar. In the accompanying figure if ABC is a triangle and DE is parallel to BC, and the lengths of the sides are as labeled, you should be able to calculate DE.



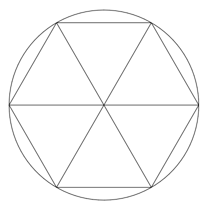
The Pythagorean Theorem: If ABC is a right triangle with the right angle at B, then



$$(AB)^2 + (BC)^2 = (AC)^2. \tag{C.117}$$

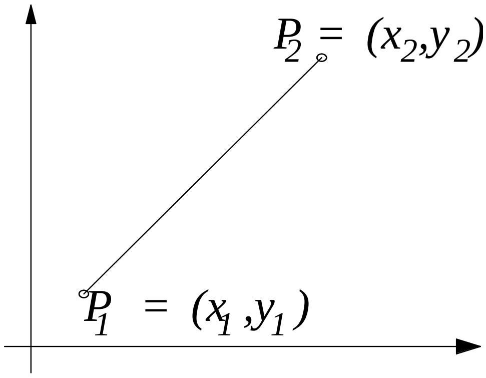
In a given circle, equal arcs subtend equal chords.

A regular hexagon inscribed in a circle has each of its sides equal to the radius of the circle. The radii joining the vertices of this hexagon to the center of the circle decompose the hexagon into six equilateral triangles.



It is assumed that you are familiar with the process of representing points in the plane by pairs of numbers. If $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are points with $x_1 \neq x_2$, then the slope of the segment joining P_1 to P_2 is defined to be

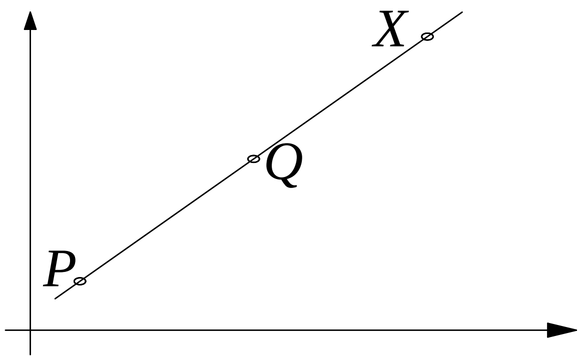
$$\text{slope}(P_1P_2) = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_1 - y_2}{x_1 - x_2} = \text{slope}(P_2P_1). \tag{C.118}$$



If $x_1 = x_2$ we say that P_1P_2 has *undefined slope*, or that P_1P_2 is a *vertical* segment. If slope(P_1P_2) is zero we say that P_1P_2 is a *horizontal* segment.

Let $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ be two distinct points in the plane. If $p_1 = q_1$, then the line passing through P and Q is defined to be the set of all points of the form (p_1, y), where y can be an arbitrary real number. If $p_1 \neq q_1$, then the *line joining P to Q* is defined to be the set consisting of P together with all points $X = (x, y)$ such that slope$(PX) =$ slope(PQ). Thus if $p_1 \neq q_1$ then (x, y) is on the line joining P to Q if and only if

$$(x,y) = (p_1,p_2) \qquad \text{or} \qquad \frac{y-p_2}{x-p_1} = \frac{q_2-p_2}{q_1-p_1}. \tag{C.119}$$



If $m =$ slope(PQ) then equations C.119 can be rewritten as

$$y = p_2 + m(x - p_1). \tag{C.120}$$

If $p_1 \neq q_1$ then (C.120) is called an *equation for the line joining P and Q*. If $p_1 = q_1$ then

$$x = p_1 \tag{C.121}$$

is called an equation for the line joining P and Q. Thus a point X is on the line joining P to Q if and only if the coordinates of X satisfy an equation for the line.

Two lines are *parallel*, (i.e. they do not intersect or they are identical,) if and only if they both have the same slope or they both have undefined slopes.

Appendix D

Some Maple Commands

In this appendix we will indicate the syntax for Maple commands corresponding to some of the ideas discussed in the notes.

Detailed descriptions and examples of any of these commands can be found by entering Maple or Xmaple and typing `help(`*command-name*`);` , e.g. `help(sum);` or `help(limit);` . If this produces a syntax error, try putting the command name in back-quotes (‘), e.g. `help(‘@‘);` .

π	`Pi;`
∞	`infinity;`
x^y	`x^y;`
$\sum_{j=a}^{b} f(j)$	`sum(f(j),j=a..b);`
$\int f(x)dx$	`int(f(x),x);`
$\int_a^b f(x)dx$	`int(f(x),x=a..b);`
$f'(x)$	`diff(f(x),x);`
$f^{(n)}(x)$	`diff(f(x),x\$n);`
f'	`D(f);`
$f^{(n)}$	`(D@@n)(f)`

$\lim_{x \to a} f(x)$	`limit(f(x),x=a);`
$\lim_{x \to a^+} f(x)$	`limit(f(x),x=a,right);`
$\lim_{x \to a^-} f(x)$	`limit(f(x),x=a,left);`
$f \circ g$	`f@g;`

Some useful Maple commands are

`%`	A name for the last line calculated by Maple.
`simplify(`*expression*`);`	simplify *expression.*
`factor(`*expression*`);`	factor *expression.*
`evalf(`*expression*`);`	express *expression* as a number in decimal notation. (Here *expression* should represent a number.)
`quit;`	Exit from Maple.

EXAMPLES

```
> sum(j^2,j=1..n);
```

$$\frac{1}{3}(n+1)^3 - \frac{1}{2}(n+1)^2 + \frac{1}{6}n + \frac{1}{6}$$

```
> simplify(%);
```

$$\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$$

```
> factor(%);
```

$$\frac{1}{6}n(n+1)(2n+1)$$

```
> limit( (n^2 + 2*n + 3)/(4*n^2 + 5*n - 1), n=infinity);
```

$$\frac{1}{4}$$

```
> 4*int( 1/(1+x^2),x=0..1);
```

$$\pi$$

```
> evalf(%);
```

$$3.141592654$$

```
> D(f@g);
```

$$\mathrm{D}(f)@g\,\mathrm{D}(g)$$

```
> limit( exp(1/x),x=0,left);
```

$$0$$

```
> limit( exp(1/x),x=0);
```

$$\textit{undefined}$$

To solve the system of equations $x^2 + y^2 = 25$, $x + y = 1$ for the unknowns x and y.

```
> solve({x^2 + y^2 = 25, x+y=1},{x,y});
```

$$\{y = -3, x = 4\}, \{y = 4, x = -3\}$$

To introduce a simple name for a complicated expression.

```
> f1 := 3*x^4 + 5*x^2 + 3;
```

$$f1 := 3\,x^4 + 5\,x^2 + 3$$

```
> diff(f1,x);
```

$$12\,x^3 + 10\,x$$

```
> diff(f1,x$2);
```

$$36\,x^2 + 10$$

To introduce a simple name for a function.

```
> g1 := (x -> (x+1)/(x-1));
```

$$g1 := x \to \frac{1+x}{x-1}$$

```
> g1(2);
```

$$3$$

```
> (g1@g1)(t);
```

$$\frac{1 + \dfrac{1+t}{-1+t}}{\dfrac{1+t}{-1+t} - 1}$$

```
> simplify(%);
```

$$t$$

In Xmaple you can draw graphs of functions by using the `plot` command. If `f` is a function, and `a`, `b`, `c`, `d` are numbers with `a` $<$ `b` and `c` $<$ `d`, then the command

`plot(f(x),x=a..b,c..d)`
will cause the part of the graph of `f` in the box B(`a,b`:`c,d`) to be drawn. The command
`> plot( x^3/(1-x^2), x=-4..4, -4..4 );`
makes the plot in figure a.

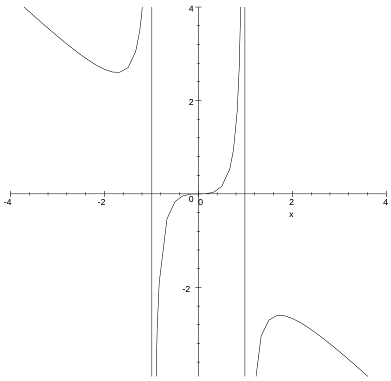


Figure a.

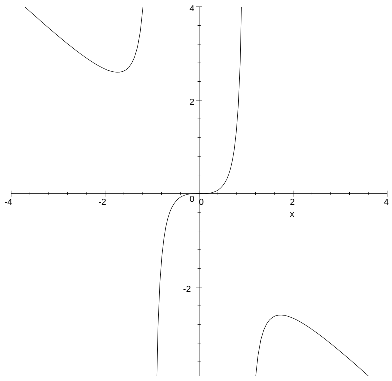


Figure b.

The two vertical lines appearing in the plot are not part of the graph, but are due to the fact that the points ± 1 are not in the the domain of the function being plotted. We can get a better plot of the graph by adding the statement `discont=true` to the plot command, to warn Maple that the function being plotted is discontinuous. (Actually, according to our definitions, this function is not discontinuous, because it is continuous at every point of its domain.) The command
`> plot( x^3/(1-x^2), x=-4..4, -4..4, discont=true);`
makes the plot in figure b.

Appendix E

List of Symbols

Index