A Summary of Calculus

Karl Heinz Dovermann

Professor of Mathematics
University of Hawaii

July 28, 2003

This publication was typeset using $\mathcal{AMS}$-TEX, the American Mathematical Society's TEX macro system, and LATEX 2_ε. The graphics were produced with the help of *Mathematica*[1].

This is an incomplete draft which will undergo further changes.

[1] *Mathematica* Version 2.2, Wolfram Research, Inc., Champaign, Illinois (1993).

Contents

Preface

In these notes we like to summarize calculus.

Chapter 1

Basic Concepts

Introduction

In this chapter we introduce limits and derivatives. These are basic concepts of calculus. We provide some rules for their computations.

1.1 Real Numbers and Functions

We assume that the reader is familiar with the real numbers (denoted by $\mathbb{R}$) and the operations of addition and multiplication. A real number is either positive, negative, or zero. This allows us to order the real numbers. If x and y are real numbers, then x is larger than y (i.e., $x > y$) if $x - y$ is positive.

Until further notice, we will work with real valued functions in one real variable. Their *domains*, the sets on which these functions are defined, are subsets of the real numbers, and they take values in $\mathbb{R}$. The *range* of a function is a set in which the function takes values. The *image* of a function f consists of all those points y in the range for which there exists an x in the domain of f, such that $f(x) = y$.

We will make frequent use of the *absolute value function*.

$$|x| = \begin{cases} x & \text{if } x > 0 \\ -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

The *distance* between two points a and b on the real line is $|a - b|$, and $\{x \in \mathbb{R} \mid |x - a| < \epsilon\}$ is the set of all real numbers whose distance from

a is less than ϵ. Expressed as an interval this set is $(a-\epsilon, a+\epsilon)$. For computations with absolute values it is worth noting that, for any two real numbers x and x

$$|x \cdot y| = |x| \cdot |y|, \quad |x+y| \leq |x| + |y|, \text{ and } ||x| - |y|| \leq |x-y|. \tag{1.1}$$

The first inequality is referred to as *triangle inequality*, and the last one is a variation of it.

Every now and then we will allude to the *completeness* of the real line, which means that every bounded subset of the real line has a least upper bound. This property is crucial for calculus, but arguments using it are too difficult for an introductory course on the subject.

1.2 Limits

Limits are a central tool in calculus and other areas of mathematics. We discuss them in this section.

Definition 1.1. *Let f be a function and L a real number. We say that*

$$L = \lim_{x \to a} f(x) \tag{1.2}$$

if for all $\epsilon > 0$ there exists a $\delta > 0$, such that $|f(x) - L| < \epsilon$ whenever x is in the domain of f and $0 < |x-a| < \delta$.

The equation in (1.2) reads as L is the *limit* of $f(x)$ as x approaches a. We also say that $f(x)$ *approaches* or *converges to* L as x approaches a. An intuitive interpretation is that the *expected value* of $f(x)$ at $x = a$ is L, based on the values of $f(x)$ for x near a.

In all but a few degenerate cases, limits are unique if they exist.

Proposition 1.2. *Suppose that $f(x)$ has a limit at $x = a$, then this limit is unique, provided that the domain of the function f contains points arbitrarily close to a.*[1] [2]

The latter assumption in the proposition is satisfied if the domain of f contains an interval, and either a belongs to this interval or a is an end point

[1] Expressed in mathematical language this means, for all $\delta > 0$ there is a point b in the domain of f, such that $0 < |b-a| < \delta$.

[2] Some authors do not apply the concept of a limit at isolated points of the domain of a function, points for which there are no other arbitrarily close points in the domain of the function.

of it. To avoid intricate language, we make this kind of an assumption for the remainder of this section. Taking limits is compatible with the basic algebraic operations in the following sense.

Proposition 1.3. *Assume that the domains of the functions $f(x)$ and $g(x)$ both contain an interval of the form (d, a) or (a, e) where $d < a < e$. Suppose that*

$$\lim_{x\to a} f(x) = L \quad \textit{and} \quad \lim_{x\to a} g(x) = M.$$

and that c is a constant. Then

$$\begin{aligned} \lim_{x\to a}(f+g)(x) &= M + L \\ \lim_{x\to a} cf(x) &= cM \\ \lim_{x\to a}(f\cdot g)(x) &= M\cdot L \\ \lim_{x\to a}(f/g)(x) &= M/L \textit{ provided that } L \neq 0. \end{aligned}$$

As a special case we obtain the following useful observation:

$$\lim_{x\to c} f(x) = L \quad \text{if and only if} \quad \lim_{x\to c}(f(x) - L) = 0. \tag{1.3}$$

Proposition 1.4 (Pinching Theorem). *Assume that the domains of the functions $f(x)$, $g(x)$, and $h(x)$ all contain an interval of the form (d, a) or (a, e) where $d < a < e$ and that $f(x) \leq h(x) \leq g(x)$. If*

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

then the limit of $h(x)$ exists as x approaches a, and it is equal to L.

For many functions the computation of limits is no challenge.

Proposition 1.5. *If $f(x)$ is a polynomial, a rational function, or a trigonometric function and $f(a)$ is defined, then*

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

The following limits are important in the calculations of some derivatives.

$$\lim_{x\to 0}\frac{\cos x - 1}{x} = 0, \quad \lim_{x\to 0}\frac{\sin x}{x} = 1. \quad \text{and} \quad \lim_{x\to a}\frac{x^n - a^n}{x - a} = na^{n-1}. \tag{1.4}$$

Hints: The first two limits follow easily from the estimates in Theorem 1.7, discussed in the following subsection. The last assertion can be proved using synthetic division, at least if n is an integer.

1.2.1 Two important estimates

In preparation of the proof of Theorem 1.7 we show

Theorem 1.6. *If* $h \in [-\pi/4, \pi/4]$, *then*

$$|\sin h| \le |h| \le |\tan h|. \tag{1.5}$$

Proof. In Figure 1.1 you see part of the unit circle. For $h \in [-\pi/4, \pi/4]$ we set $C = (\cos h, \sin h)$. Given two points X and Y in the plane, the distance between them is denoted by $\overline{XY}$. We denote by $\widehat{BC}$ the length of the arc (part of the unit circle) between B and C.

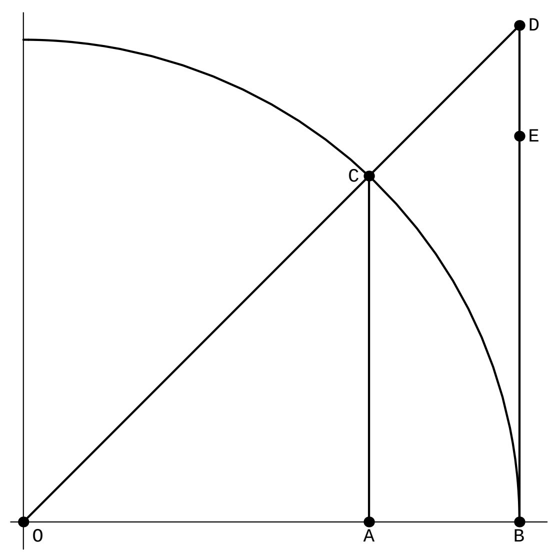

Figure 1.1: The unit circle

We find that $|\sin h| = \overline{AC} \le |h| = \widehat{BC}$ because going from C straight down to the x-axis is shorter than following the circle from C to the x-axis.

Secondly, to show that $|h| = \widehat{BC} \le |\tan h| = \overline{BD}$, imagine that you roll the circle along the vertical line through B until the point C touches it in the point E. We use the process of rolling the circle along the line to measure $|h|$. In particular, $|h| = \overline{BE}$. It appears to be clear[3] that $\overline{BE} \le \overline{BD}$. This

[3] Here our argument relies on intuition. A rigorous argument requires work. One can show that the area of a disk with radius one is π. From this is follows by elementary geometry that the area of the slice of the disk with vertices O, B and C has area $|h|/2$. This slice is contained in the triangle with vertices O, B and D, and the area of the slice is $(\tan|h|)/2$. It follows that $|h| \le \tan|h|$.

verifies that $|h| \le \tan|h|$, the second inequality which we claimed in the theorem. □

Theorem 1.7. *If* $h \in [-\pi/4, \pi/4]$, *then*[4]

$$|1-\cos h| \le \frac{h^2}{2} \quad \textit{and} \quad |h-\sin h| \le \frac{h^2}{2}. \tag{1.6}$$

Proof of Theorem 1.7. In Figure 1.2 you see half of a circle of radius 1 centered at the origin, and a triangle with vertices A, B, and C. Let $h \in [-\pi/4, \pi/4]$ be the number for which we want to show the inequality and $C = (\cos h, \sin h)$. Denote by $\overline{XY}$ the length of the straight line segment between the points X and Y. Let $\widehat{BC}$ be the length of the arc (part of the unit circle) between B and C.

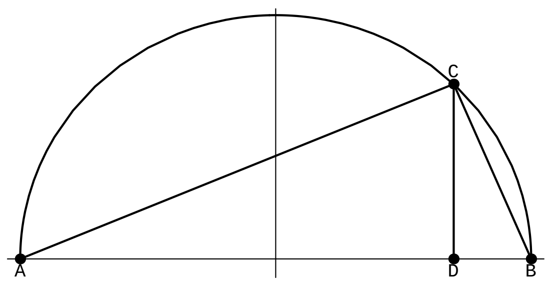

Figure 1.2: The unit circle

From the picture we read off that

$$\overline{AB} = 2, \quad \overline{DB} = (1-\cos h), \quad \widehat{BC} = |h|, \text{ and } \quad \overline{BC} \le \widehat{BC}.$$

Using similar triangles we see $\overline{AB}/\overline{BC} = \overline{BC}/\overline{DB}$ and $(\overline{BC})^2 = \overline{AB} \times \overline{DB}$. In other words

$$2(1-\cos h) = \overline{AB} \times \overline{DB} = (\overline{BC})^2 \le (\widehat{BC})^2 = h^2.$$

[4]The inequalities hold without the restriction on h, but we only need them on an interval around zero. Restricting ourselves to this interval simplifies the proofs somewhat.

The first estimate in (1.6) is an immediate consequence.

If $h = 0$, then both sides of the second inequality in (1.6) are zero, verifying the assertion in this case. If $0 \neq h \in [-\pi/4, \pi/4]$, then Theorem 1.6 tells us that

$$|\sin h| \leq |h| \leq |\tan h| = \frac{|\sin h|}{\cos h} \quad \text{hence} \quad 0 \leq \cos h \leq \frac{\sin h}{h} \leq 1.$$

Subtracting the terms in this inequality from 1 we find

$$0 \leq 1 - \frac{\sin h}{h} \leq 1 - \cos h \leq 1.$$

Using our previous estimate for $|1 - \cos h|$ and our assumption that $|h| \leq \pi/4 < 1$, we conclude that

$$\left|\frac{h - \sin h}{h}\right| \leq |1 - \cos h| \leq \frac{h^2}{2} \leq \frac{h}{2}.$$

The second estimate claimed in the theorem is an immediate consequence. $\square$

1.3 More Limits

The material in the previous, first section about limits suffices for a while. In some situations one would like to modify the definition in Section 1.2, and we do so in this section. The first two limits express how the function behaves as we approach a point a from the right or left. They are called the *right* and *left hand* limits. The next two limits express what happens as the variable tends to plus or minus infinity. We call them *limits at infinity.* The last two limits allow us to express that the values of a function tend to plus or minus infinity. We call them *infinite limits.*

Definition 1.8. *Let f be a function and L a real number. We say that*

$$L = \lim_{x \to a^+} f(x)$$

if for all $\epsilon > 0$ there exists a $\delta > 0$, such that $|f(x) - L| < \epsilon$ whenever x is in the domain of f and $a < x < a + \delta$.

Definition 1.9. *Let f be a function and L a real number. We say that*

$$L = \lim_{x \to a^-} f(x)$$

if for all $\epsilon > 0$ there exists a $\delta > 0$, such that $|f(x) - L| < \epsilon$ whenever x is in the domain of f and $a - \delta < x < a$.

For example, if $f(x) = \text{sign}(x) = x/|x|$, then

$$\lim_{x \to a^+} f(x) = 1 \quad \text{and} \quad \lim_{x \to a^-} f(x) = -1.$$

We can consider what happens to the values of a function $f(x)$ as x approaches ∞ or $-\infty$.

Definition 1.10. *Let f be a function and L a real number. We say that*

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

if for all $\epsilon > 0$ there exists a number M, such that $|f(x) - L| < \epsilon$ whenever x is in the domain of f and $x > M$.

Definition 1.11. *Let f be a function and L a real number. We say that*

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

if for all $\epsilon > 0$ there exists a number M, such that $|f(x) - L| < \epsilon$ whenever x is in the domain of f and $x < M$.

For example

$$\lim_{x \to \infty} \frac{1}{x} = 0 \quad \text{and} \quad \lim_{x \to -\infty} \frac{1}{1 + x^2} = 1.$$

Definition 1.12. *Let f be a function and a a real number. We say that*

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

if for all M there exists a $\delta > 0$ such that $f(x) > M$ whenever x is in the domain of f and $0 < |a - x| < \delta$.

In other words, we can make sure that the value of $f(x)$ is larger than any given number M, no matter how large, by taking x close to a.

Definition 1.13. *Let f be a function and a a real number. We say that*

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

if for all M there exists a $\delta > 0$ such that $f(x) < M$ whenever x is in the domain of f and $0 < |a - x| < \delta$.

In the last two definitions a may be replaced by $a^{\pm}$, so that we approach a from the left or right, and a can be replaced by $\pm\infty$.

For example

$$\lim_{x \to 0^+} \frac{1}{x} = \infty \quad \text{and} \lim_{x \to \infty} \sqrt{x} = \infty.$$

1.4 Continuous Functions

We define continuous functions and discuss a few of their basic properties. The class of continuous functions will play a central role later.

Definition 1.14. *Let f be a function and c a point in its domain. The function is said to be continuous at c if for all $\epsilon > 0$ there exists a $\delta > 0$, such that $|f(c) - f(x)| < \epsilon$ whenever x belongs to the domain of f and $|x - c| < \delta$. A function f is continuous if it is continuous at all points in its domain.*

In most cases the condition in Definition 1.14 says that

$$\lim_{x \to c} f(x) = f(c). \tag{1.7}$$

In fact, this equation holds whenever there are points in the domain of f arbitrarily close to c. See the footnote to Proposition 1.2. If c is an isolated point in the domain of f, i.e., there are no other points in the domain of f arbitrarily close to c, then the function is always continuous at c.

Polynomials, rational functions, and trigonometric functions are continuous. One can produce many more continuous functions through standard operations on functions.

Proposition 1.15. *Let f and g be continuous functions. Then $f + g$, $f \cdot g$, f/g and $f \circ g$ are continuous, wherever these functions are defined.*

The clarify the remark about the domain in the proposition, we note that the function $(f + g)(x) = f(x) + g(x)$ is defined for those x for which both f and g are defined. The same statement holds for $(f \cdot g)(x) = f(x) \cdot g(x)$. To determine the domain of f/g one needs to exclude those points where g is zero. For the composition $(f \circ g)(x) = f(g(x))$ on needs that g takes values in the domain of f.

One may also reverse the order of applying a continuous function and calculating a limit:

$$\lim_{x \to c} f(g(x)) = f\left(\lim_{x \to c} g(x)\right), \tag{1.8}$$

provided the natural technical assumption hold, i.e., g is defined at points arbitrarily close to c, f is defined for all $g(x)$ where x is in the domain of g and close to c, and f is continuous at $\lim_{x \to c} g(x)$.

Theorem 1.16 (Intermediate Value Theorem). *Suppose that f is defined and continuous on the closed interval $[a, b]$. If C is in between $f(a)$ and $f(b)$, then there exists a $c \in [a, b]$, such that $f(c) = 0$.*

E.g., suppose that $p(x) = x^3 - x^2 + 2x - 1$. The polynomial is certainly a continuous function, $p(0) = -1$ and $p(1) = 1$. According to the theorem there exists some $c \in (0, 1)$, such that $p(c) = 0$.

Theorem 1.17 (Extreme Value Theorem). *Let f be defined and continuous on the closed interval $[a, b]$. Then there exist points c and d in $[a, b]$, such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [a, b]$.*

Expressed in words, the theorem says that a continuous function on a closed interval assumes a smallest and largest value.

The Intermediate Value and Extreme Value theorem are typically proved in an introductory analysis course. They are equivalent to the completeness of the real line. We mentioned this property of the real numbers in Section 1.1.

1.5 Lines

In general, a line consists of the points (x, y) in the plane which satisfy the equation

$$ax + by = c \tag{1.9}$$

for some given real numbers a, b and c, where it is assumed that a and b are not both zero. The line is vertical if and only if $b = 0$. If $b \neq 0$ we may rewrite the equation as

$$y = -\frac{a}{b}x + \frac{c}{b} = mx + B. \tag{1.10}$$

The number m is called the *slope* of the line, and B is the point in which the line intersects the y-axis, also called the *y-intercept*. Given any two points (x_1, y_1) and (x_2, y_2) in the plane, the line through them has slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

For our purposes, the most useful version of the equation of a line is its *point-slope formula*. The equation of a line with slope m through the point (x_1, y_1) is

$$y = m(x - x_1) + y_1. \tag{1.11}$$

1.6 Tangent Lines and the Derivative

We like to introduce the concept of tangent lines. To be able to express ourselves concisely, let us say

Definition 1.18. *A point c is an* interior point *of a subset B of $\mathbb{R}$ if there is an open interval I, such that $c \in I \subseteq B$.*

We give a first definition for a tangent line.

Definition 1.19. *Suppose $f(x)$ is a function and c is an interior point of its domain. We call a line $t(x)$ the* tangent line *to the graph of $f(x)$ at $x = c$ if $t(x)$ is the best linear approximation of $f(x)$ on some open interval around c, i.e., the line $t(x)$ is closer to the graph of $f(x)$ than any other line for all x in some open interval around c.*

For a given function and an interior point c in its domain there may or may not be a tangent line, but it there is a tangent line, then it is unique.

Although the term *'best linear approximation near c'* gives an excellent intuitive picture what a tangent line is, this definition is hard to work with. It is easier to work with a more concrete definition.

Definition 1.20. *Suppose $f(x)$ is a function and c is an interior point of its domain. We call a line $t(x)$ the* tangent line *to the graph of $f(x)$ at $x = c$ if*

$$\lim_{x \to c} \frac{f(x) - t(x)}{x - c} = 0. \tag{1.12}$$

The equation in (1.12) expresses in a precise form in which sense the tangent line is close to the graph of $f(x)$ near c. Not only does $f(x) - t(x)$ converge to zero as x approaches c, it does so even when divided by $x - c$.

We use tangent lines to define the concept of differentiability and the derivative.

Definition 1.21. *Suppose $f(x)$ is a function and c is an interior point of its domain, and assume that there is a tangent line to the graph of $f(x)$ at $x = c$. Then we say that $f(x)$ is* differentiable *at c. We call the slope of the tangent line the* derivative *of $f(x)$ at c, and we denote it by $f'(c)$.*

Utilizing the notation in the previous definition we can write down the equation of the tangent line to the graph of $f(x)$ at $x = c$ in point-slope form:

$$t(x) = f'(c)(x - c) + f(c). \tag{1.13}$$

To *differentiate* a function means to find its derivative.

By definition, an *open* set is a set, such that each of its points is an interior point.

Definition 1.22. *Suppose the domain of the function $f(x)$ is an open set. Then say that $f(x)$ is differentiable if it is differentiable at each point of its domain. We consider $f'(x)$ as a function, whose domain consists of all those points where $f(x)$ is differentiable.*

Example 1.23. Let $p(x) = 2x^4 - 3x^2 + 5$. Find the tangent line $t(x)$ to the graph of $p(x)$ at $x = -2$ and $p'(-2)$.

Solution: As a first step we expand p in powers of $u = (x + 2)$. To do so, we substitute $u - 2$ for x and expand p in powers of u. You are expected to fill in some of the arithmetic steps.

$$\begin{aligned} p &= 2(u-2)^4 - 3(u-2)^2 + 5 \\ &= 2(u^4 - 8u^3 + 24u^2 - 32u + 16) - 3(u^2 - 4u + 4) + 5 \\ &= 2u^4 - 16u^3 + 45u^2 - 52u + 25 \end{aligned}$$

Reversing the substitution, replacing u by $(x + 2)$, we find:

$$p(x) = 2(x+2)^4 - 16(x+2)^3 + 45(x+2)^2 - 52(x+2) + 25.$$

We assert that $t(x) = -52(x + 2) + 25$ and $p'(-2) = -52$.

For $t(x)$ as proposed, we see that

$$\begin{aligned} \left|\frac{p(x)-t(x)}{x-c}\right| &= \left|\frac{p(x)-t(x)}{x+2}\right| \\ &= |2(x+2)^3 - 16(x+2)^2 + 45(x+2)| \\ &\le 65|x+2| \quad (\text{provided } |x+2| \le 1) \end{aligned}$$

This estimate shows that $(p(x) - t(x))/(x - c)$ converges to zero as x approaches $c = -2$. By definition, this means that $t(x)$ is the desired tangent line. Its slope is $p'(-2) = -52$. ◊

The example is generic. We can use any polynomial $p(x)$ and point $x = c$ and write $p(x)$ in powers of $(x - c)$. Say, the result is

$$p(x) = A_n(x-c)^n + \cdots + A_1(x-c) + A_0.$$

The technique used in the example, suitably generalized, shows that

$$t(x) = A_1(x-c) + A_0$$

is the tangent line to the graph of $p(x)$ at $x = c$, and $p'(c) = A_1$. Eventually we will find a more efficient method for differentiating polynomials, but we have shown that

Proposition 1.24. *Polynomials are differentiable.*

1.6.1 Derivatives without Limits

Without a doubt, the definition of a limit is the most difficult one in a first semester of calculus, and it is interesting to explore ways to develop calculus, rigorously, without the limit concept. One can do this by replacing the condition in (1.12) by a slightly stronger one.

Definition 1.25. *Suppose $f(x)$ is a function and c is an interior point of its domain. We call a line $t(x)$ the* tangent line *to the graph of $f(x)$ at $x = c$ if there exists and open interval I around c and a number A, such that*

$$|f(x) - t(x)| \le A(x - c)^2 \tag{1.14}$$

for all $x \in I$.

With this definition fewer functions will be differentiable than with the one given in Definition 1.20, but this is not crucial.

The inequality in (1.14) can be rewritten as

$$q(x) = t(x) - A(x - c)^2 \le f(x) \le t(x) + A(x - c)^2 = p(x),$$

where the parabolas $q(x)$ and $p(x)$ are defined by the expressions they are adjacent to. All four function $f(x)$, $q(x)$, $p(x)$, and $t(x)$ have the same value at $x = c$. In an example, this situation is shown in Figure 1.3. There you see the function $f(x) = \sin x$, the parabola $p(x)$ (dotted and open upwards), the parabola $q(x)$ (dotted and open downwards), and the tangent line $t(x)$ (dashed). The parabolas $p(x)$ and $q(x)$ touch each other without crossing, and the picture shows how they 'hug' each other. There is very little space left between $p(x)$ and $q(x)$, and $f(x)$ and $t(x)$ are squeezed in between them. In this sense, the graphs of $f(x)$ and $t(x)$ have to be close to each other near $x = c$.

A pedagogical advantage of the approach is that one does not have to understand limits before one can understand the definition of the derivative. There is also a geometric picture which illustrates the concept of *closeness*, *tangent line*, and *derivative*. The condition in (1.14) is also more accessible to computer assisted algebra than the limit definition. In terms of algebraic geometry (1.14) at least alludes to a divisibility condition.

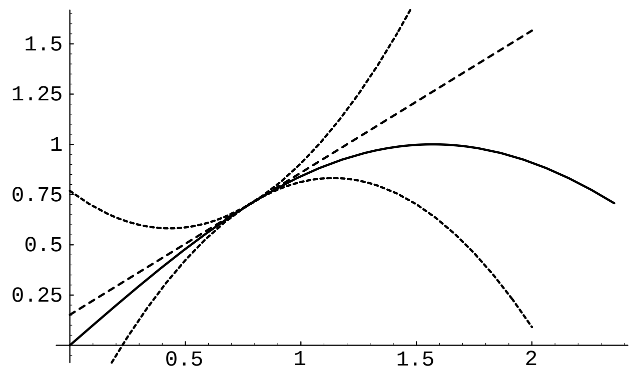

Figure 1.3: Sine Function and Tangent Line between two Parabolas

1.7 Secant Lines and the Derivative

Often a different approach is taken to motivate and introduce the derivative.

Theorem 1.26. *Suppose f is a function and c is an interior point of its domain. If f is differentiable at c, then*

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

Proof. This is obvious once one uses the expression for the tangent line in (1.13) and substitutes it in the expression in (1.12) inside the limit.

$$\frac{f(x) - t(x)}{x - c} = \frac{f(x) - f(c)}{x - c} - f'(c). \tag{1.15}$$

Apply limits to both sides of the equation and the assertion follows. ☐

Let us explain the situation geometrically. Suppose a and b are distinct points in the domain of the function f. The line through $(a, f(a))$ and $(b, f(b))$ is called a *secant line*, and its slope $(f(a) - f(b))/(a - b)$ is called the *average rate of change* of f over the interval $[a, b]$. In (1.15) we are considering the slopes of secant lines through $(c, (f(c))$ and $(x, f(x))$, and then we take the limit as x approaches c. The theorem asserts that for a differentiable function this limit of the slopes of secant lines is the slope of the tangent line. For the obvious reason $f'(c)$ is called the *rate of change* or *instantaneous rate of change* of f at c.

Many authors introduce the derivative as the limit of the slopes of secant lines, call $t(x) = f'(x - c) + f(c)$ the tangent line, and possibly illustrate that the tangent line is close to the graph in the sense of Definition 1.20.

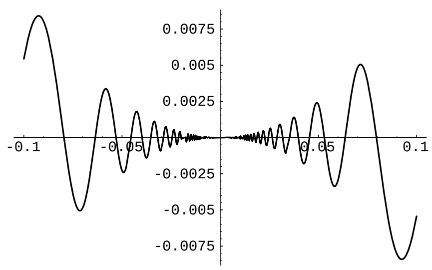

Figure 1.4: $f(x) = x^2 \sin(1/x)$

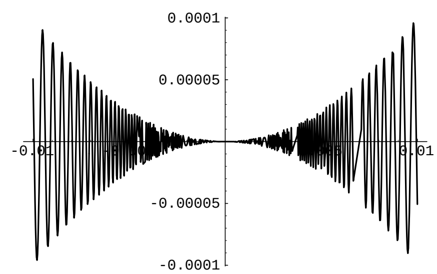

Figure 1.5: $f(x) = x^2 \sin(1/x)$

It is misleading to say that the graph of $f(x)$ looks like, or resembles, a line near c. Eventually you will be able to show that the function

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

is differentiable everywhere on the real line. You see part of its graph over two different intervals in Figure 1.4 and 1.5. By no stretch of imagination will you say that the graph of the function looks like a line.

1.8 Differentiability implies Continuity

It is worth pointing out that

Theorem 1.27. *If a function is differentiable at a point, then it is continuous at this point.*

Proof. Denote the function by $f(x)$ and the point of differentiability by c. By assumption we have the derivative $f'(c)$ and

$$\lim_{x \to c} \left[\frac{f(x) - f(c)}{x - c} - f'(c) \right] = 0.$$

Then certainly

$$\lim_{x \to c} [(f(x) - f(c)) - f'(c)(x - c)] = 0.$$

Because $f'(c)(x-c)$ converges to zero as x approaches c, so does $(f(x)-f(c))$. This implies that $\lim_{x\to c} f(x) = f(c)$ and that $f(x)$ is continuous at c. □

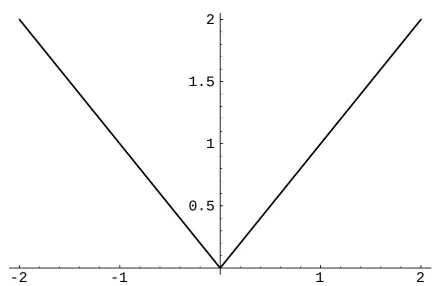

Figure 1.6: The absolute value function

The converse of the theorem is false. There are continuous functions which are not differentiable. E.g., the function $f(x) = |x|$ is continuous, but it is not differentiable at $x = 0$. It is apparent from the graph (see Figure 1.6) that there is not line close to the graph of this function near $x = 0$.

We can also give an analytic argument. According to the definition of differentiability, we have to study the difference quotients $(|x|-|0|)/(x-0) = |x|/x$. They are 1 if $x > 0$ and -1 if $x < 0$. There is no number these difference quotients converge to, and $f(x) = |x|$ is not differentiable at $x = 0$.

1.9 Basic Examples of Derivatives

Let us use the definitions and work out a few derivatives.

Example 1.28. If $f(x) = x^n$ and n is a non-negative integer, i.e., $n = 0$, 1, 2, … , then $f'(x) = nx^{n-1}$.

Proof. Suppose that $n \geq 2$. Then

$$\lim_{x \to c} \frac{x^n - c^n}{x - c} = \lim_{x \to c} (x^{n-1} + x^{n-2}c + \cdots xc^{n-2} + c^{n-1}) = nc^{n-1}$$

The cases $n = 0$ and $n = 1$ are even easier and left to the reader. □

Example 1.29. If $f(x) = 1/x$, then $f'(x) = -1/x^2$.

Proof. Suppose $c \neq 0$.

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

□

Example 1.30. If $f(x) = \sqrt{x}$ and $x > 0$, then $f'(x) = 1/(2\sqrt{x})$.

Proof.

$$\lim_{x \to c} \frac{\sqrt{x} - \sqrt{c}}{x - c} = \lim_{x \to c} \frac{x - c}{(x - c)(\sqrt{x} + \sqrt{c})} = \frac{1}{2\sqrt{c}}.$$

□

Remark 1. Eventually we will see that if $f(x) = x^a$ for any real number a, then $f'(x) = ax^{a-1}$, generalizing all of the examples above.

Exercise 1. *Suppose that* $f(x) = \sqrt{ax + b}$ *and* $ax + b > 0$. *Show that*

$$f'(x) = \frac{a}{2\sqrt{ax + b}}.$$

The tangent line to the graph of $f(x)$ *at* $x = c$ *is then*

$$t(x) = \frac{a}{2\sqrt{ac + b}}(x - c) + \sqrt{ac + b}.$$

Verify that

$$|f(x) - t(x)| \leq \frac{a^2}{2(\sqrt{ac + b})^3}(x - c)^2. \tag{1.16}$$

The estimate in (1.16) shows differentiability in the sense of Definition 1.25, and provides an explicit error estimate, a bound on the difference between the function and its tangent line.

Example 1.31. Show that $\sin' x = \cos x$. For this equation to hold, the angle x needs to be measured in radians.

Proof. Below we will set $x = c + h$ and $x - c = h$.

$$\begin{aligned}
\lim_{x\to c}\left[\frac{\sin x - \sin c}{x - c}\right] &= \lim_{h\to 0}\left[\frac{\sin(c+h) - \sin c}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\sin c\cos h + \cos c\sin h - \sin c}{h}\right] \\
&= \lim_{h\to 0}\left[\frac{\sin c(\cos h - 1) + \cos c\sin h}{h}\right] \\
&= \sin c \cdot \lim_{h\to 0}\frac{\cos h - 1}{h} + \cos c \cdot \lim_{h\to 0}\frac{\sin h}{h} \\
&= \cos c.
\end{aligned}$$

For computation of the limits in the second to last line see (1.4). □

The tangent line to the graph of the sine function at $x = c$ is

$$t(x) = \cos c(x - c) + \sin c.$$

It is left as an exercise for the reader to show that

$$|\sin x - t(x)| \leq (x - c)^2 \tag{1.17}$$

The steps are essentially the same as in the proof above. The estimate in (1.17) does not only show differentiability in the sense of Definition 1.25, but it provides an explicit error estimate, a bound on the difference between the function and its tangent line.

Exercise 2. *If $f(x) = \cos x$, then $f'(x) = -\sin x$. The details are similar to the ones in Example 1.31. Furthermore, if*

$$t(x) = \sin c(x - c) + \cos x$$

is the tangent line to the graph of $f(x)$ at $x = c$, then

$$|f(x) - t(x)| \leq (x - c)^2.$$

1.10 The Exponential and Logarithm Functions

The exponential and logarithm are of great importance and we do not want to delay their introduction any further. Still, technically we are not quite prepared for it and at a later point we have to revisit the introduction to fill in details.

Suppose a is a positive real number and $a \neq 1$. For any rational number $r = p/q$ (p and q are integers) one can define $a^r = \sqrt[q]{a^p}$. First we take a p-th power and then a q-root. In this sense we have a function $h(r) = a^r$, whose domain consists of all rational numbers. This function is monotonic. More precisely, $h(r)$ is increasing if $a > 1$ and decreasing when $0 < a < 1$.

Theorem-Definition 1.32. *Let a be a positive number, $a \neq 1$. There exists exactly one monotonic function, called the* exponential function with base a *and denoted by* $\exp_a(x)$*, which is defined for all real numbers x such that* $\exp_a(x) = a^x$ *whenever x is a rational number. Furthermore, $a^x > 0$ for all x, so that the domain of the exponential function is $(-\infty, \infty)$. For every number $y > 0$ there exists exactly one number x, such that $\exp_a(x) = y$, so we use $(0, \infty)$ as the range of the exponential function $\exp_a(x)$.*

It is common, and we will follow this convention, to use the notation a^x for $\exp_a(x)$ also if x is not rational. The arithmetic properties of the exponential function, also called the exponential laws, are collected in our next theorem. The theorem just says that the exponential laws, which you previously learned for rational exponents, also hold in the generality of our current discussion.

Theorem 1.33 (Exponential Laws). *For any positive real number a and all real numbers x and y*

$$\begin{aligned} a^x a^y &= a^{x+y} \\ a^x / a^y &= a^{x-y} \\ (a^x)^y &= a^{xy} \end{aligned}$$

If x is the unique solution of the equation $a^x = y$, then we set

$$\log_a(y) = x. \tag{1.18}$$

We just defined a function $\log_a(y)$. It is called the *logarithm function* with base a, and by construction it is the inverse of the exponential function $\exp_a(x)$. More explicitly,

$$a^{\log_a y} = y \quad \text{and} \quad \log_a(a^x) = x$$

for all $x \in \mathbb{R}$ and all $y > 0$. The domain of the logarithm function is $(0, \infty)$ and its range is $(-\infty, \infty)$. It is increasing if $a > 1$ and decreasing if $0 < a < 1$.

Corresponding to the exponential laws in Theorem 1.33 we have the laws of logarithms. One set of laws implies the other one, and vice versa.

Theorem 1.34 (Laws of Logarithms). *For any positive real number $a \neq 1$, for all positive real numbers x and y, and any real number z*

$$\begin{aligned} \log_a(xy) &= \log_a(x) + \log_a(y) \\ \log_a(x/y) &= \log_a(x) - \log_a(y) \\ \log_a(x^z) &= z \log_a(x) \end{aligned}$$

In Figures 1.7 and 1.8 you see parts of the graphs of the exponential and logarithm functions with base 2.

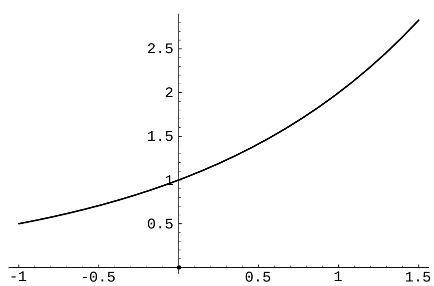

Figure 1.7: $\exp_2(x)$

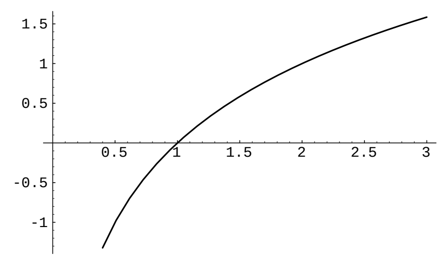

Figure 1.8: $\log_2(x)$

The Euler number e as base

There is one number which is preferrable as base over the others. This irrational number is called the *Euler number* (named after Leonard Euler) and denoted by e, and $e \approx 2.718281828$. We will define it precisely later.

Definition 1.35. *The* exponential function *is the exponential function for the base e. It is denoted by* $\exp(x)$ *or* e^x. *Its inverse is the* natural logarithm function. *It is denoted by* $\ln(x)$. *So* $\exp(x) = \exp_e(x)$ *and* $\ln(x) = \log_e(x)$.

Eventually we will see

$$\exp'(x) = \exp(x) \quad \text{and} \quad \ln'(x) = \frac{1}{x}. \tag{1.19}$$

The derivative of the exponential function is the exponential function, and the derivative of the natural logarithm function is $1/x$.

Other Bases

Finally, let us relate the exponential and logarithm functions for different bases to those with base e. For any positive number a ($a \neq 1$),

Theorem 1.36.

$$a^x = e^{x \ln a} \quad \textit{and} \quad \log_a x = \frac{\ln x}{\ln a}.$$

These identities follow from the exponential laws and the laws of logarithms.

1.11 Differentiability on Closed Intervals

In Definition 1.22 we defined what it means that a function is differentiable on an open set. There are situations in which one would like to apply the notion of differentiability to functions with other kinds of domains. Let us formalize the idea of extending functions.

Definition 1.37. *Suppose that I and J are subsets of the real line $\mathbb{R}$ and $I \subseteq J$, that I is the domain of a function f, and that J is the domain of a function F. We call F an* extention *of f if it agrees with f on I, i.e., $F(x) = f(x)$ for all $x \in I$.*

Definition 1.38. *A function f is said to be differentiable on a subset I of $\mathbb{R}$ if it extends to a differentiable function F on an open set. We set $f'(x) = F'(x)$ for all $x \in I$.*

Without some restrictions on I, a function may be differentiable without the derivative being well defined. The least technical and for our purposes sufficient solution is captured in

Proposition 1.39. *Suppose the function f is defined on an interval I, the interval is neither empty nor a single point, and f extends to a differentiable function F on an open interval containing I, then $f'(x) = F'(x)$ is unique for all $x \in I$.*

We are mostly concerned with defining differentiability for functions whose domain is a closed interval $[a, b]$, where $a < b$. Some authors use one-sided limits and one-sided derivatives to contemplate derivatives at the end points of the interval. Our discussion is less painful, and it lends itself more to generalizations in higher dimensions.

Let us discuss two examples. The function $f(x) = x^2$ with domain $[0, 1]$ is differentiable. It extends to the differentiable function $F(x) = x^2$ with the open set $(-\infty, \infty)$ as its domain. In contrast, the function $g(x) = \sqrt{x}$ is not differentiable on the interval $[0, \infty)$. The only sensible candidate for the tangent line to the graph of $g(x)$ at the point $(0, 0)$ is a vertical line. The slope of this line is not a real number and we do not have a derivative. (The function $g(x)$ is differentiable if we use $(0, \infty)$ as domain.)

1.12 Other Notations for the Derivative

There are different notations for the derivative of a function. Physicists will indicate a derivative with respect to time by a dot. E.g., if x is a function of time, then they will write $\dot{x}(t)$ instead of $x'(t)$. Leibnitz' notation for the derivative of a function f of a variable x is $\frac{df}{dx}$. We will use it frequently. Expressing the derivatives of the exponential and natural logarithm functions this way (see (1.19)) we have:

$$\text{If } y(x) = e^x, \text{ then } \frac{dy}{dx} = y = e^x, \text{ and if } y(x) = \ln x, \text{ then } \frac{dy}{dx} = \frac{1}{x}.$$

This notation is not always specific enough. The expression dy/dx stands for the derivative of y with respect to x, and that is a function. The expression does not tell where dy/dx is evaluated. To be specific about this aspect, it makes sense to write (compare Example 1.31):

$$\text{If } y(x) = \sin x, \text{ then } \frac{dy}{dx}(x) = \cos x.$$

In this notation x plays two roles. It is the name of the variable of y as well as the name of the variable of the derivative of y. This in acceptable because it won't lead to confusion. Instead of $\frac{df}{dx}(x)$ we also write $\frac{d}{dx} f(x)$. This is particularly convenient if f stands for a larger expression as in

$$\frac{d}{dx} \sin x = \cos x \quad \text{or} \quad \frac{d}{dx} e^x = e^x.$$

1.13 Rules of Differentiation

We discuss formulas for calculating the derivative of a composite function from the derivatives of its constituents. These formulas, together with the knowledge of the derivatives of some basic functions, turn the process of differentiation for many functions into an algorithm, a rather mechanical process. You can do it even on the computer, which means that no "understanding" is required. You are expected to learn the basic rules, be able to apply the accurately, and practice many examples. In the last section of this chapter we summarize the computational results of this section. We collect the rules established in this section and tabulate the derivatives of many of the important functions which we considered.

1.13.1 Linearity of the Derivative

Differentiation is compatible with addition of functions and multiplication with a constant. In a more mathematical language one says that differentiation is *linear*. Let f and g be functions, and assume that both of them are differentiable at x. Let c be a real number. Then $f+g$ and cf are differentiable at x and their derivatives are given by

$$(f+g)'(x) = f'(x) + g'(x) \quad \text{and} \quad (cf)'(x) = cf'(x). \tag{1.20}$$

In Leibnitz' notation this reads

$$\frac{d}{dx}(f+g)(x) = \frac{df}{dx}(x) + \frac{dg}{dx}(x) \quad \text{and} \quad \frac{d}{dx}(cf)(x) = c\frac{df}{dx}(x). \tag{1.21}$$

In words, the derivative of a sum of functions is the sum of the derivatives, and the derivative of a multiple of a function is the multiple of the derivative.

Example 1.40. Differentiate

$$h(x) = x^2 + 3e^x.$$

Solution: Set $f(x) = x^2$, $g(x) = e^x$ and $c = 3$. Then $h(x) = f(x) + 3g(x)$. Previously we found that $f'(x) = 2x$ and that $g'(x) = e^x$, see (1.19). We conclude that

$$h'(x) = \frac{dh}{dx}(x) = 2x + 3e^x. \quad \Diamond$$

Example 1.41. Differentiate $\log_a x$, the logarithm functions for an arbitrary positive base a, $a \neq 1$.

Solution: Recall that $\log_a x = \frac{\ln x}{\ln a}$, see Theorem 1.36. In this sense $\log_a x = cf(x)$ where $c = 1/\ln a$ and $f(x) = \ln x$. We stated previously that $\ln' x = 1/x$, see (1.19). Using the linearity of the derivative, we find

$$\log_a' x = \frac{d}{dx}\left(\frac{\ln x}{\ln a}\right) = \frac{1}{\ln a}\ln' x = \frac{1}{\ln a} \times \frac{1}{x} = \frac{1}{x \ln a}. \quad \diamondsuit$$

Suppose f and g are defined and differentiable on an set. Thinking of f and g more as functions, and not so much as functions evaluated at a point, we may omit (x) from the notation. Then the differentiation rules are

$$(f+g)' = f' + g' \quad \text{or} \quad \frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx} \tag{1.22}$$

and

$$(cf)' = cf' \quad \text{or} \quad \frac{d}{dx}(cf) = c\frac{df}{dx}. \tag{1.23}$$

Example 1.42. Find the derivative of an arbitrary polynomial.

Solution: A polynomial is a finite sum of multiples of non-negative powers of the variable, i.e., a function of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where the a_i are constants. Using Example 1.28 and the linearity of the derivative we see right away that

$$f'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1.$$

Here is a specific example, a special case of the formula which we just derived.

$$\text{If } f(x) = 4x^5 - 3x^2 + 4x + 5, \text{ then } \quad f'(x) = 20x^4 - 6x + 4. \quad \diamondsuit$$

1.13.2 Product and Quotient Rules

Next we state the *product* and the *quotient rule*. They allow us to calculate the derivatives of products and quotients of functions. Again, let f and g be functions, and assume that both of them are differentiable at x. For the quotient rule assume in addition that $g(x) \neq 0$. Then the product fg and the quotient f/g are differentiable at x and their derivatives are given by

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x) \tag{1.24}$$

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}. \tag{1.25}$$

In Leibnitz' notation these formulas become

$$\frac{d}{dx}(fg)(x) = \frac{df}{dx}(x)g(x) + f(x)\frac{dg}{dx}(x) \tag{1.26}$$

$$\frac{d}{dx}\left(\frac{f}{g}\right)(x) = \frac{\frac{df}{dx}(x)g(x) - f(x)\frac{dg}{dx}(x)}{[g(x)]^2}. \tag{1.27}$$

Example 1.43. Differentiate the function $h(x) = x^2 \ln x$.

Solution: Write $h(x) = f(x)g(x)$ with $f(x) = x^2$ and $g(x) = \ln x$. Then $f'(x) = 2x$ and $g'(x) = 1/x$, see (1.19). Putting this into the product formula yields

$$h'(x) = f'(x)g(x) + f(x)g'(x) = 2x \ln x + x^2 \frac{1}{x} = x(2 \ln x + 1). \quad \diamondsuit$$

Example 1.44. Find the derivative of the rational function.

$$r(x) = \frac{x^2 - 5}{x^3 + 1}.$$

Solution: We set $p(x) = x^2 - 5$ and $q(x) = x^3 + 2$. Then $p'(x) = 2x$ and $q'(x) = 3x^2$. According to the quotient rule

$$r'(x) = \frac{2x(x^3 + 1) - (x^2 - 5)3x^2}{(x^3 + 1)^2} = \frac{-x^4 + 15x^2 + 2x}{(x^3 + 1)^2}. \quad \diamondsuit$$

Example 1.45. The formula

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

for all integer powers n. If $n \leq -1$, then we domain of the function is $\mathbb{R}\backslash\{0\}$, the real line with the origin removed.

Solution: We verified this formula for $n \geq 0$ in Example 1.28. Let n be a negative integer and $m = -n$. Then

$$\frac{d}{dx}x^n = \frac{d}{dx}\left[\frac{1}{x^m}\right] = \frac{0 \cdot x^m - 1 \cdot mx^{m-1}}{x^{2m}} = \frac{-m}{x^{m+1}} = nx^{n-1}.$$

Example 1.46. Find the derivative of

$$f(x) = \tan x.$$

Solution: We express $f(x)$ as a quotient of two functions, $f(x) = \sin x / \cos x$, and apply the quotient rule. Use that $\sin' x = \cos x$ (see Example 1.31) and $\cos' x = -\sin x$ (see Exercise 2 on page 17). We find

(1.28)
$$\tan' x = \frac{\sin' x \cos x - \sin x \cos' x}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x.$$

Some books and computer programs will give this result in a different form. Based on the relevant trigonometric identity, they write

$$\tan' x = 1 + \tan^2 x. \tag{1.29}$$

That draws our attention to the fact that the function $f(x) = \tan x$ satisfies the differential equation

$$f'(x) = 1 + f^2(x). \quad \Diamond$$

Example 1.47. Differentiate the function

$$f(x) = \sec x.$$

Solution: We write the function as a quotient: $f(x) = 1/\cos x$. The function is defined for all x for which $\cos x \neq 0$, i.e., for x not of the form $n\pi + 1/2$, where n is an integer. We apply the quotient rule, using that $\cos' x = -\sin x$ (see Exercise 2 on page 17), and that the derivative of a constant vanishes. We find

$$\sec' x = \frac{\sin x}{\cos^2 x} = \frac{\sin x}{\cos x} \cdot \frac{1}{\cos x} = \tan x \sec x. \quad \Diamond \tag{1.30}$$

Suppose f and g are defined and differentiable on an open set. Thinking of f and g again more as functions, and not so much as functions evaluated at a point, we may once more omit (x) from the notation. Then the product rule and quotient rule become

$$(fg)' = f'g + fg' \quad \text{or} \quad \frac{d}{dx}(fg) = \frac{df}{dx}g + f\frac{dg}{dx} \tag{1.31}$$

and, wherever $g(x) \neq 0$,

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \quad \text{or} \quad \frac{d}{dx}\left(\frac{f}{g}\right) = \frac{\frac{df}{dx}g - f\frac{dg}{dx}}{g^2}. \tag{1.32}$$

Here g^2 is the square of the function g, given by $g^2(x) = [g(x)]^2$.

1.13.3 Chain Rule

Let f and g be functions, and suppose that the domain of f contains the range of g, so that the composition $(f \circ g)(x) = f(g(x))$ is defined for all x in the domain of g. Set $h = f \circ g$, so that $h(x) = f(g(x))$. The *chain rule* says that whenever g is differentiable at x and f is differentiable at $g(x)$, then $h(x)$ is differentiable at x and

$$h'(x) = (f \circ g)'(x) = f'(g(x))g'(x). \tag{1.33}$$

In Leibnitz' notation the chain rule says that

$$\frac{dh}{dx}(x) = \frac{d}{dx}f(g(x)) = \frac{df}{du}(g(x))\frac{dg}{dx}(x). \tag{1.34}$$

Example 1.48. Differentiate the function

$$h(x) = e^{x^2+1}.$$

Solution: We write $h = f \circ g$ as a composition of two functions, with $g(x) = x^2 + 1$ and $f(u) = e^u$. Remember that $f'(u) = f(u) = e^u$ and $g'(x) = 2x$. In particular, $f'(g(x)) = e^{x^2+1}$. The chain rule tells us that

$$h'(x) = f'(g(x))g'(x) = 2xe^{x^2+1}.$$

In the last expression we reversed the order of the factors to make the expression more readable. ♢

Example 1.49. Let $u(x)$ be a differentiable function.

$$\text{If} \quad f(x) = e^{u(x)} \quad \text{then} \quad f'(x) = u'(x)e^{u(x)}.$$

Here are some specific examples:

$$\begin{aligned}
\frac{d}{dx}e^{2x+5} &= 2\ e^{2x+5} \\
\frac{d}{dx}e^{\sin x} &= \cos x\ e^{\sin x} \\
\frac{d}{dx}e^{\tan x} &= \sec^2 x\ e^{\tan x}. \quad \diamondsuit
\end{aligned}$$

Example 1.50. Combining Example 1.45 with the chain rule we find

$$\frac{d}{dx}u^n(x) = nu'(x)u^{n-1}(x)$$

for all integers n, assuming only that u is differentiable at x and $u(x) \neq 0$ if $n \leq -1$.

Solution: Set $g(x) = u(x)$ and $f(u) = u^n$. Then

$$h(x) = f(g(x)) = u^n(x).$$

According to Example 1.45, $f'(u) = nu^{n-1}$. The chain rule tells us now that

$$h'(x) = f'(g(x))g'(x) = n(g(x))^{n-1}g'(x) = nu'(x)u^{n-1}(x).$$

We reordered the expressions so that the expression is more readable.

To be specific, here are concrete examples:

$$\begin{aligned}
\frac{d}{dx}(3x+5)^8 &= 8(3x+5)^{8-1} \cdot 3 = 24(3x+5)^7 \\
\frac{d}{dx}(x^2+1)^{25} &= 25(x^2+1)^{24} \cdot 2x = 50x(x^2+1)^{24} \\
\frac{d}{dx}\tan^3 x &= 3\sec^2 x \tan^2 x \\
\frac{d}{dx}\cos^2 x &= 2\cos x(-\sin x) = -2\cos x \sin x \\
\frac{d}{dx}\sec^5 x &= 5\sec^4 x \sec x \tan x = 5\sec^5 x \tan x. \quad \diamondsuit
\end{aligned}$$

Example 1.51. Differentiate the function $\ln|u|$ for $u \neq 0$.

Solution: We asserted that $\ln' u = 1/u$ for positive values of u, see (1.19). So, suppose that $u < 0$. Then $u = -|u|$ and $\ln|u| = \ln(-u)$. The chain rule tells us that, for $u < 0$,

$$\frac{d}{du}\ln|u| = \frac{1}{|u|}\frac{d}{du}(-u) = (-1)\frac{1}{-u} = \frac{1}{u}.$$

This means that for all non-zero u

$$\frac{d}{du}\ln|u| = \frac{1}{u}. \quad \diamondsuit \tag{1.35}$$

More generally, envoking the chain rule

$$\frac{d}{dx}\ln|u(x)| = \frac{u'(x)}{u(x)}, \tag{1.36}$$

assuming that u is differentiable and nowhere zero on its domain. E.g.,

$$\frac{d}{dx}\ln|x^2-4| = \frac{2x}{x^2-4}$$

for all $x \neq \pm 2$.

We push matters a bit further. We use the formulae for differentiating the exponential and natural logarithm functions. Eventually we will verify them independently.

Consider a function u which is differentiable and nowhere zero on its domain and q any real number. Then

$$\text{(1.37)} \qquad \text{If} \quad f(x) = |u(x)|^q \quad \text{then} \quad f'(x) = q\frac{u'(x)}{u(x)}|u(x)|^q.$$

The assertion follows from (1.36), Example 1.49 and the exponential laws.

$$\begin{aligned} f'(x) &= \frac{d}{dx} e^{\ln f(x)} \\ &= \frac{d}{dx} e^{\ln(|u(x)|^q)} \\ &= \frac{d}{dx} e^{q \ln |u(x)|} \\ &= \left[\frac{d}{dx}\left(q \ln |u(x)|\right)\right] e^{q \ln |u(x)|} \\ &= q\frac{u'(x)}{u(x)}|u(x)|^q. \end{aligned}$$

Here is a concrete example:

$$\frac{d}{dx}\left|\frac{1}{2} - \sin x\right|^5 = 5\frac{-\cos x}{\frac{1}{2} - \sin x}\left|\frac{1}{2} - \sin x\right|^5$$

whenever $\sin x \neq 1/2$. Specifically, we have to exclude all x of the form $\frac{\pi}{6} + 2n\pi$ and $\frac{5\pi}{6} + 2n\pi$, where n is an arbitrary integer.

For differentiable functions which are everywhere positive on their domain and any real number q the differentiation formula in (1.37) specializes to

$$\text{(1.38)} \qquad \frac{d}{dx} u^q(x) = qu'(x)u^{q-1}(x).$$

For example:

$$\begin{aligned} \frac{d}{dx}(\sin x)^{1/2} &= \frac{\cos x}{2\sqrt{\sin x}} \quad \text{for } x \in (0, \pi) \text{ and} \\ \frac{d}{dx}(\sec^2 x + 5)^{\pi} &= 2\pi \sec^2 x \tan x (\sec^2 x + 5)^{\pi - 1} \quad \text{for } x \in (-\pi/2, \pi/2). \end{aligned}$$

Using the tricks from above, we get the following derivatives:

$$\begin{aligned}
\frac{d}{dx}a^x &= a^x \ln a \quad (\text{Assume } a > 0. \text{ Hint: } a^x = e^{x \ln a}) \\
\frac{d}{dx}x^x &= (1 + \ln x)x^x \quad (\text{Assume } x > 0,\ x \neq 1. \text{ Hint: } x^x = e^{x \ln x}) \\
\frac{d}{dx}x^{\sin x} &= \left(\frac{\sin x}{x} + \cos x \ln x\right) x^{\sin x} \quad (\text{Assume } x \in (0, \pi/4)).
\end{aligned}$$

To differentiate a composition of more than two differentiable functions we apply the chain rule repeatedly. E.g.,

$$\frac{d}{dx}f(g(h(x))) = f'(g(h(x))\frac{d}{dx}g(h(x)) = f'(g(h(x)) \cdot g'(h(x)) \cdot h'(x).$$

For example

$$\begin{aligned}
\frac{d}{dx}e^{\sqrt{x^2+1}} &= e^{\sqrt{x^2+1}} \cdot \frac{1}{2\sqrt{x^2+1}} \cdot 2x = \frac{xe^{\sqrt{x^2+1}}}{\sqrt{x^2+1}} \\
\frac{d}{dx}\tan^3(5x^2 - x + 5) &= 3\tan^2(5x^2 - x + 5)\sec^2(5x^2 - x + 5) \cdot (10x - 1)
\end{aligned}$$

1.13.4 Hyperbolic Functions

The exponential function may be used to define the *hyperbolic sine* and *cosine.*

$$\sinh x = \frac{1}{2}\left[e^x - e^{-x}\right] \quad \& \quad \cosh x = \frac{1}{2}\left[e^x + e^{-x}\right] \tag{1.39}$$

You are invited to verify that

$$\cosh^2 x - \sinh^2 x = 1.$$

Conversely, one can show that any point (u, v) on the hyperbola

$$u^2 - v^2 = 1$$

can be expressed as $(\pm\cosh x, \sinh x)$ for some $x \in (-\infty, \infty)$. These observations motive the attribute ‘hyperbolic’.

It is elementary to compute the derivatives of the hyperbolic functions:

$$\sinh' x = \cosh x \quad \text{and} \quad \cosh' x = \sinh x.$$

One may also define other hyperbolic functions

$$\tanh x = \frac{\sinh x}{\cosh x}, \ \coth x = \frac{\cosh x}{\sinh x}, \ \operatorname{sech} x = \frac{1}{\cosh x}, \ \text{and} \ \operatorname{csch} x = \frac{1}{\sinh x}.$$

As a routine application of the rules of differentiation, you may calculate the derivatives of these functions. There are identities for these hyperbolic functions, comparable to the identities for the trigonometric functions. You can find them in any table of mathematical formulas, or you can work them out yourself.

1.13.5 Derivatives of Inverse Functions

Let us recall. Two functions f and g are said to be *inverses* of each other (or each function is the inverse of the other one) if the domain of f is equal to the range of g, the domain of g is equal to the range of f, and

$$g(f(x)) = x \quad \text{and} \quad f(g(y)) = y \tag{1.40}$$

for all x in the domain of f and all y in the domain of g. A few essential properties of inverse functions are listed in

Proposition 1.52. *Suppose f and g are inverses of each other.*

1. *The graph of g is obtained from the graph of f by reflection at the diagonal.*

2. *If f is increasing, then so is g. If f is decreasing, then so is g.*

3. *If f is continuous, then it is monotonic (increasing or decreasing) on any interval in its domain.*

4. *If f is continuous and I is an interval in the domain of f, then $J = f(I)$, the image of I under the map f, is an interval. If I is an open interval, then J is an open interval.*

Some parts of this proposition are elementary, others are consequences of the intermediate value theorem.

For example, the function $f(x) = \cos x$ maps the interval $[0, \pi]$ to the inteval $[-1, 1]$. The function $f(x) = e^x$ maps the interval $(-\infty, \infty)$ to the interval $(0, \infty)$. The function $\tan x$ maps the interval $(-\pi/2, \pi/2)$ to the interval $(-\infty, \infty)$. It is customary to define its inverse $\arctan x$ as a function from $(-\infty, \infty)$ to $(-\pi/2, \pi/2)$. The function $\sin x$ maps the interval $[-\pi/2, \pi/2]$ to the interval $[-1, 1]$. Its inverse $\arcsin x$ is typically used with

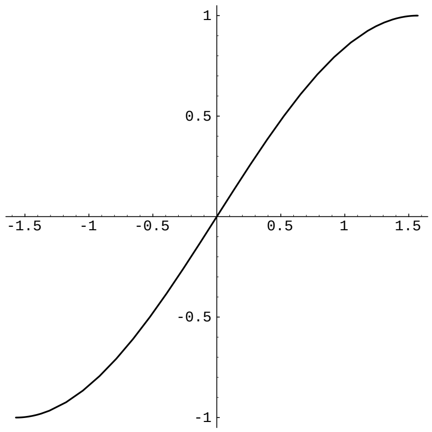

Figure 1.9: $\sin x$ on $[-\pi/2, \pi/2]$

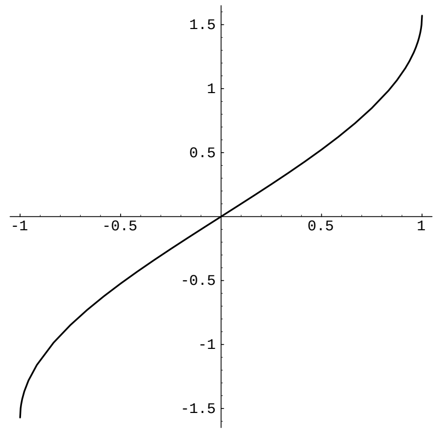

Figure 1.10: $\arcsin y$ on $[-1, 1]$

domain $[-1, 1]$, and its range is $[-\pi/2, \pi/2]$. You see the graph of these two functions in Figures 1.9 and 1.10.

The relation between the derivative of a function and its inverse is spelled out in our next theorem.

Theorem 1.53. *Let f be a differentiable and invertible function which is defined on an open interval (a, b), and denote the image of f by (A, B). Denote the inverse of f by g. Then g is differentiable at all points $y \in (A, B)$ for which $f'(g(y)) \neq 0$. For these values of y and for x such that $f(x) = y$ the derivative is given by:*

$$g'(y) = \frac{1}{f'(g(y))} \quad \textit{or} \quad g'(f(x)) = \frac{1}{f'(x)}.$$

Proof. We will not give a formally complete proof of the differentiability assertion. Still, if the line $t(x)$ is close to the graph of the function $f(x)$ at the point $(x, f(x))$ and $y = f(x)$, then its reflection $T(x)$ at the diagonal is close to the graph of the function $g(x)$ at the point $(f(x), x) = (y, g(y))$. We need that $T(x)$ is not vertical, and this is assured by the assumption that $t(x)$ is not horizontal. With the role of x and y being interchanged, the slope of $t(x)$ is the reciprocal of the slope of $T(x)$. This provides the formula for the derivative. Actually, this is also easy to calculate.

By definition we have $f(g(y)) = y$ for all $y \in (A, B)$. Differentiate both sides of the equation. We find

$$f'(g(y))g'(y) = 1 \quad \text{and} \quad g'(y) = \frac{1}{f'(g(y))},$$

as claimed. If $y = f(x)$, then $g(y) = g(f(x)) = x$, and we obtain the second version of the formula for the derivative of the inverse of the function:

$$g'(f(x)) = \frac{1}{f'(x)}.$$

□

We apply the theorem to find some important derivatives.

Example 1.54. Assume that the natural logarithm function is differentiable and that $\ln' x = 1/x$, as asserted in (1.19). Show that the exponential function is differentiable and that

$$\frac{d}{dy}e^y = e^y.$$

Solution: By definition, the exponential function is the inverse of the natural logarithm function ln. Set $f(x) = \ln x$ and $g(y) = e^y$ in Theorem 1.53. We note that $\ln'(x) \neq 0$ for all x in $(0, \infty)$, the domain of the natural logarithm. The theorem says that the exponential function is differentiable and provides the formula for the derivative:

$$\frac{d}{dy}e^y = \frac{1}{\ln'(e^y)} = \frac{1}{1/e^y} = e^y,$$

as claimed. ◊

Example 1.55. Show that the function $g(y) = \arctan y$ (the inverse of $f(x) = \tan x$) is differentiable, and that

$$\frac{d}{dy}\arctan y = \frac{1}{1 + y^2}.$$

According to standard conventions we use $(-\infty, \infty)$ as the domain and $(-\pi/2, \pi/2)$ as the range for arctan.

Solution: The function $f(x) = \tan x$ is differentiable on its entire domain, and $f'(x) = \sec^2 x$ is nowhere zero. Theorem 1.53 tells us that $g(y) = \arctan y$ is differentiable on its entire domain $(-\infty, \infty)$. The theorem also provides us with the formula for the derivative:

$$\arctan'(y) = \frac{1}{\tan'(\arctan y)} = \frac{1}{\sec^2(\arctan y)} = \cos^2(\arctan y).$$

All we need to do now is to figure out what $\cos^2(\arctan y)$ is. To do this we draw a triangle in which we identify the available data. We refer to the notation in Figure 1.11.

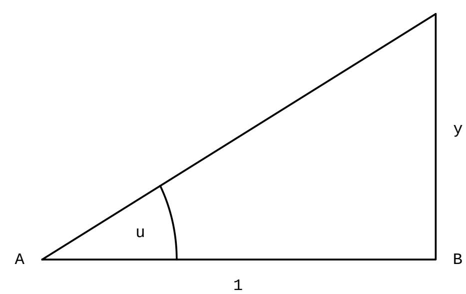

Figure 1.11: An informative triangle

There you see a rectangular triangle, the right angle is at the vertex B. The angle at the vertex A is called u. The adjacent side to this angle is chosen to be of length 1, and the opposing side of length y. So, by definition,

$$\tan u = y \quad \text{and} \quad \arctan y = u.$$

By the theorem of Pythagoras, the length of the hypotenuse is $\sqrt{1+y^2}$. Then

$$\cos u = \frac{1}{\sqrt{1+y^2}} \quad \text{and} \quad \cos^2(\arctan y) = \frac{1}{1+y^2}.$$

The conclusion is that

$$\arctan'(y) = \frac{1}{1+y^2}. \tag{1.41}$$

This is exactly what we claimed. $\Diamond$

Combined with the chain rule, and assuming the differentiability of $u(x)$, we find a slightly more general formula:

$$\frac{d}{dx}\arctan(u(x)) = \frac{u'(x)}{1+u^2(x)}. \tag{1.42}$$

For example:

$$\begin{aligned} \frac{d}{dx}\arctan(x^2+5) &= \frac{2x}{1+(x^2+5)^2} \\ \frac{d}{dx}\arctan(\sin x) &= \frac{\cos x}{1+\sin^2 x}. \end{aligned}$$

The reader is invited to verify the formulas for the other inverse trigonometric functions $\arcsin x$, $\arccos x$, $\operatorname{arccot} x$, and $\operatorname{arcsec} x$ as they are given in Table 1.3 on page 63. For example

Exercise 3. *It is customary to think of* $\arcsin x$ *as a function from* $[-1, 1]$ *to* $[-\pi/2, \pi/2]$*. Show that* $\arcsin x$ *is differentiable on* $(-1, 1)$*, and that its derivative is*

$$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}.$$

We may once more improve on this formula. Let $u(x)$ be a differentiable function which is defined on an open interval, and suppose that $|u(x)| < 1$. Then, using the chain rule, we find that

$$\frac{d}{dx}\arcsin(u(x)) = \frac{u'(x)}{\sqrt{1-u^2(x)}}. \tag{1.43}$$

For example:

$$\begin{aligned} \frac{d}{dx}\arcsin(3x) &= \frac{3}{\sqrt{1-9x^2}} \quad \text{if } x \in (-1/3, 1/3) \\ \frac{d}{dx}\arcsin(x^2) &= \frac{2x}{\sqrt{1-x^4}} \quad \text{if } x \in (-1, 1) \end{aligned}$$

1.13.6 Implicit Differentiation

Until now we considered functions which were given explicitly. I.e., we were given an equation $y = f(x)$, where $f(x)$ is some instruction which assigns a value to x. The points on the graph of f are the points which satisfy the equation. Consider the equation

$$(x^2+y^2)^2 = x^2 - y^2. \tag{1.44}$$

The solutions of this equation form a curve[5] in the plane called a lemniscate, see Figure 1.12. Parts of this curve look like the graph of a function, such

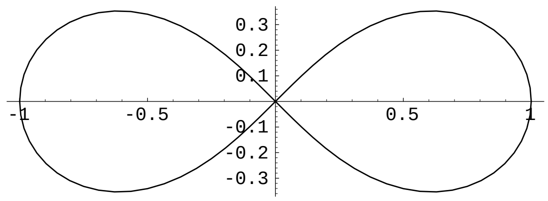

Figure 1.12: Lemniscate

as the points for which $y \geq 0$. Without solving the equation for y, we still like to calculate the slope of curve at one of its points. This process is called *implicit differentiation*.

Let us start out with an example which we have studied before.

Example 1.56. The unit circle consists of all points which satisfy the equation $x^2 + y^2 = 1$. Find the slope of the tangent line to the unit circle at the point $(1/2, \sqrt{3}/2)$.

Solution: We write $y = y(x)$ to emphasize that y as a function of x. Differentiating both sides of the equation of the circle we get

$$2x + 2y\frac{dy}{dx} = 0 \quad \text{or} \quad \frac{dy}{dx} = \frac{-x}{y}.$$

Plugging in the coordinates of the specified point, we find that

$$\left.\frac{dy}{dx}\right|_{(1/2,\sqrt{3}/2)} = \frac{-1}{\sqrt{3}}.$$

We used a different way to indicate at which point we evaluate the derivative because we had to specify the x and the y coordinate of the point. ♢

[5]We will rely on the readers intuitive idea of a curve in the plane.

Example 1.57. Find the slope of the tangent line to the lemniscate

$$(x^2 + y^2)^2 = x^2 - y^2,$$

and find the coordinates of the points where the tangent line is horizontal.

Solution: You see a picture of the lemniscate in Figure 1.12. As in Example 1.56, we consider y as a function of x and differentiate both sides of the equation. We find

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

Bring all terms with a factor dy/dx to the left hand side of the equation and those without to the right hand side.

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

Finally we get an explicit expression for $\frac{dy}{dx}$ in terms of x and y:

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

Given any point (x, y) with $y \neq 0$ on the lemniscate, we can plug it into the expression for $\frac{dy}{dx}$ and we get the slope of the curve at this point.

E.g, the point $(x, y) = (\frac{1}{2}, \frac{1}{2}\sqrt{-3 + 2\sqrt{3}})$ is a point on the lemniscate, and at this point the slope of the tangent line is

$$\frac{dy}{dx} = \frac{2 - \sqrt{2}}{\sqrt{3}\sqrt{-3 + 2\sqrt{3}}}.$$

This specific calculation takes a bit of arithmetic skill and effort to carry out.

The tangent line is horizontal whenever $\frac{dy}{dx} = 0$. A quick look at Figure 1.12 tells us that we may ignore points where $x = 0$ or $y = 0$. That means that $\frac{dy}{dx} = 0$ whenever

$$1 - 2(x^2 + y^2) = 0 \quad \text{or} \quad x^2 + y^2 = \frac{1}{2}.$$

Substitute $x^2 + y^2 = \frac{1}{2}$, and $y^2 = \frac{1}{2} - x^2$ into the equation of the curve. Then we get an equation in one variable:

$$\frac{1}{4} = x^2 - \left(\frac{1}{2} - x^2\right) \quad \text{or} \quad x^2 = \frac{3}{8} \quad \text{and} \quad y^2 = \frac{1}{8}.$$

The points at which the tangent line to the lemniscate is horizontal are

$$(x, y) = (\pm\frac{\sqrt{6}}{4}, \pm\frac{\sqrt{2}}{4}) \approx (\pm.6124, \pm.3536). \quad \diamond$$

Example 1.58. Suppose you drop a circle of radius 1 into a parabola with the equation $y = 2x^2$. At which points will the circle touch the parabola?[6]

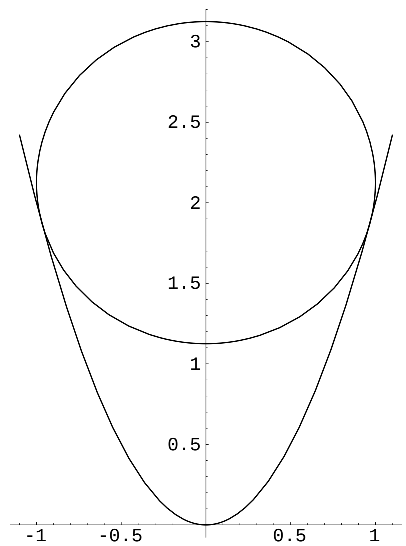

Figure 1.13: Ball in a Cup.

Solution: You see a picture of the problem in Figure 1.13. The crucial observation in this example is, that the tangent line to the parabola and the circle will be the same at the point of contact.

Suppose the coordinates of the center of the circle are $(0, a)$, then its equation is $x^2 + (y - a)^2 = 1$. Differentiating the equation of the parabola with respect to x, we find that $\frac{dy}{dx} = 4x$. Differentiating the equation of the circle with respect to x, we get

$$2x + 2(y - a)\frac{dy}{dx} = 0.$$

Assuming that $\frac{dy}{dx}$ is the same for both curves at the point of contact, we substitute $\frac{dy}{dx} = 4x$ into the second equation. After some implifications we

[6]More sensibly, drop a ball of radius 1 into a cup whose vertical cross section is the parabola $y = 2x^2$.

find:

$$x(1 + 4(y - a)) = 0.$$

The ball it too large to fit into the parabola and touch at $(0, 0)$. So we may assume that $x \neq 0$. Solving the equation $1 + 4(y - a) = 0$ for y, we find that the y coordinate of the point of contact is $y = a - \frac{1}{4}$. We substitute this expression into the equation of the circle and find that the x coordinate of the point of contact is $x = \pm\frac{\sqrt{15}}{4}$. Substituting this into the equation of the parabola, we find that $y = \frac{15}{8}$ at the point of contact. In summary, the circle touches the parabola in the points

$$(x, y) = \left(\pm\frac{\sqrt{15}}{4}, \frac{15}{8}\right). \quad \diamondsuit$$

Exercise 4. Consider the curve given by the equation

$$x^3 + y^3 = 1 + 3xy^2.$$

Find the slope of the curve at the point $(x, y) = (2, -1)$.

Exercise 5. Consider the curve given by the equation $x^2 = \sin y$. Find the slope of the curve at the point with coordinates $x = 1/\sqrt[4]{2}$ and $y = \pi/4$.

Exercise 6. Repeat Example 1.57 with the curve given by the equation $y^2 - x^2(1 - x^2) = 0$. You find a picture of this Lissajous figure in Figure 1.14.

1.14 Related Rates

Many times you encounter situations in which you have two related variables, you know at which rate one of them changes, and you like to know at which rate the other one changes. In this section we treat such problems.

Example 1.59. Suppose the radius of a ball changes at a rate of 2 cm/min. At which rate does its volume change when $r = 20$ cm?

Solution: Denote the volume of the ball by V and its radius by r. We use t to denote the time variable. We consider V as a function of r as well as t. The formula for the volume of a ball is $V(r) = \frac{4\pi}{3}r^3$. According the the chain rule:

$$\frac{dV}{dt} = \frac{dV}{dr}\frac{dr}{dt} = 4\pi r^2\frac{dr}{dt}.$$

With $r = 20$ and $\frac{dr}{dt} = 2$ we get $\frac{dV}{dt} = 3200\pi$ cm^3/min. This is the rate at which the volume of the ball changes with respect to time. $\diamondsuit$

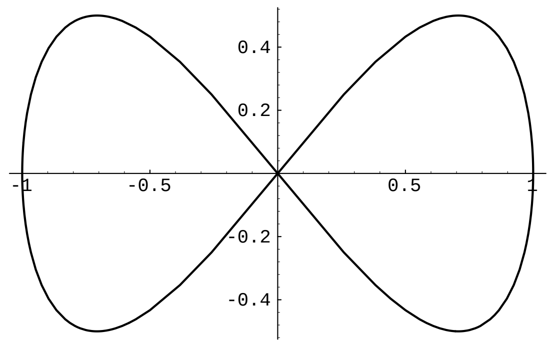

Figure 1.14: $y^2 - x^2(1 - x^2) = 0$

Example 1.60. Suppose a particle moves on a circle of radius 10 cm and centered at the origin $(0, 0)$ in the Cartesian plane. At some time the particle is at the point $(5, 5\sqrt{3})$ and moves downwards at a rate of 3 cm/min. At which rate does it move in the horizontal direction?

Solution: The equation of the circle is $x^2 + y^2 = 100$. We consider both variables, x and y, as functions of the time variable t. Implicit differentiation of the equation of the circle gives us the equation

$$2x\frac{dx}{dt} + 2y\frac{dy}{dt} = 0.$$

In the given situation $x = 5$, $y = 5\sqrt{3}$, and $\frac{dy}{dt} = -3$. We find that $\frac{dx}{dt} = 3\sqrt{3}$, so that the particle is moving to the right at a rate of $3\sqrt{3}$ cm/min. $\diamondsuit$

Example 1.61. Pressure (P) and volume (V) of air at room temperature are related by the equation[7]

$$PV^{1.4} = C.$$

[7]Boyle-Mariotte described the relation between the pressure and volume of a gas. They derived the equation $PV^\gamma = C$. It is called the adiabatic law. The constant γ depends on the molecular structure of the gas and the temperature. For the purpose of this problem, we suppose that $\gamma = 1.4$ for air at room temperature.

Here C is a constant. At some instant t_0 the pressure of the gas is 25 kg/cm^2 and the volume is 200 cm^3. Find the rate of change of P if the volume increases at a rate of 10 cm^3/min.

Solution: We consider P as a function of V. Differentiation of the equation yields

$$\frac{dP}{dV}V^{1.4} + 1.4PV^{.4} = 0 \quad \text{or} \quad \frac{dP}{dV} = -\frac{1.4P}{V}.$$

According to the chain rule

$$\frac{dP}{dt} = \frac{dP}{dV}\frac{dV}{dt} = -\frac{1.4P}{V}\frac{dV}{dt}.$$

Substituting the given information we find that the pressure decreases at a rate of 1.75 kg/cm^{2}sec. ♢

Example 1.62. The mass M of a particle at velocity v, as perceived by an observer in resting position, is

$$M(v) = \frac{m}{\sqrt{1 - v^2/c^2}},$$

where m is that mass at rest and c is the speed of light. This formula is from Einstein's special theory of relativity. At which rate is the mass changing when the particle's velocity is 90% of the speed of light, and increasing at .001c per second?

Solution: According to our rules of differentiation

$$\frac{dM}{dv} = \frac{mvc}{(c^2 - v^2)^{3/2}}.$$

Applying the chain rule and substituting the values, we find

$$\frac{dM}{dt} = \frac{dM}{dv}\frac{dv}{dt} = \frac{mvc^2}{1000(c^2 - v^2)^{3/2}} = \frac{9\sqrt{19}m}{3610} \approx .010867m.$$

The perceived mass increases at a rate of approximately 1% of its mass at rest. ♢

Exercise 7. A ladder, 7 m long, is leaning against a wall. Right now the foot of the latter is 1 m away from the wall. You are pulling the foot of the ladder further away from the wall at a rate of .1 m/sec. At which rate is the top of the ladder sliding down the wall?

1.15 Exponential Growth and Decay

An idealistic, but very useful model for population growth is the *Malthusian Law*

$$A'(t) = aA(t). \tag{1.45}$$

It says that the rate of change of a population is proportional to its size. We denoted the proportionality factor by a. We saw that the functions $A(t) = Ce^{at}$ are solutions of this equation, and it can be shown that on an interval any solution is of this form. We also say that $A(t)$ *grows exponentially* and a is the *relative growth rate*.

The equation in (1.45) is an example of a *differential equation*, an equation which involves a function and its derivatives, and the unknown is a function.

We may specify the value of A at some time t_0, say $A_0 = A(t_0)$. Then we have an *initial value problem*

$$A'(t) = aA(t) \quad \text{and} \quad A_0 = A(t_0). \tag{1.46}$$

Theorem 1.63. *On an interval which contains t_0 the function*

$$A(t) = A_0 e^{a(t-t_0)}$$

is the unique solution[8] *of the initial value problem in* (1.46).

The essential aspects of dealing with (1.46) are addressed in

Example 1.64. Suppose the size of a population of bacteria in a laboratory experiment is $C_1 = 5,000$ at time $t_1 = 2$ and $C_2 = 7,000$ at time $t_2 = 5$. Here time is measured in hours since the beginning of the experiment.

1. Find the relative growth rate a of the population.
2. Find the formula for the size of the population at any time $t \geq 0$.
3. Predict the size of the population at time $t = 10$.
4. Find the time at which the population reaches $8,000$.
5. Find the time within which the population doubles[9].

[8] That the function satisfies the differential equation follows from (1.19), which we still need to prove. The uniqueness assertion follows from Proposition 2.9 on page 68.

[9] Note that the doubling time depends only on the relative growth rate a.

Solution: We denote the size of the population at time t by $A(t)$. The theorem tells us that $A(t) = A_0 e^{a(t-t_0)}$, where $A_0 = A(t_0)$.

1. To calculate the relative growth rate a observe that

$$\frac{C_2}{C_1} = \frac{A_0 e^{a(t_2-t_0)}}{A_0 e^{a(t_1-t_0)}} = e^{a(t_2-t_1)} \quad \text{and} \quad \ln\left(\frac{C_2}{C_1}\right) = a(t_2 - t_1).$$

We find that

$$a = \frac{\ln C_2 - \ln C_1}{t_2 - t_1} = \frac{\ln 1.4}{3} \approx .11.$$

The population grows at a rate of about 11% per hour.

2. and 3. The size of the population at any time $t \geq 0$ is

$$A(t) = 5000 e^{a(t-2)},$$

where a is as above. Substituting $t = 10$ we find that $A(10) \approx 12,264$.

4. Suppose the size of the population reaches $8,000$ at time t_1, then

$$8000 = 5000 e^{a(t_1-2)} \quad \text{or} \quad \ln(8/5) = a(t_1 - 2) \quad \text{or} \quad t_1 = \frac{\ln(1.6)}{a} + 2 \approx 6.2$$

The size of the population reaches $8,000$ about 6.2 hours into the experiment.

5. Suppose at some time t_0 the size of the population is $A_0 = A(t_0)$ and T hours later the size of the population is $2A_0 = A(t_0 + T)$. Then

$$A(t_0 + T) = A_0 e^{aT} = 2A_0 \quad \text{or} \quad e^{aT} = 2 \quad \text{and} \quad aT = \ln 2.$$

Thus the doubling time is $T = \frac{\ln 2}{a} \approx 6.18$ hours.

Consider a radioactive substance. Experiments have shown that the rate at which *radioactive decays* occur is proportional to the amount of radioactive material present. This rate is proportional to the rate at which the amount of the material decreases. Suppose t denotes time and $A(t)$ the amount of radioactive substance at time t. The experience which we just described can be expressed as a differential equation

$$A'(t) = -kA(t). \tag{1.47}$$

The minus sign in the equation is included so that k will be positive. The *half-life* T of a radioactive substance is the time within which half of it decays. As in the computation of the doubling time in the previous example, one finds

$$T = \frac{\ln 2}{k}. \tag{1.48}$$

In the late 1940ies Willard Libby invented (or discovered) the method of carbon-14 dating. He was awarded the Nobel price for it. In brief, the idea is as follows. Carbon-14 occurs naturally in the atmosphere, and the amount is believed to have been essentially constant for a long time (until recent nuclear testing). All living organisms absorb it. Within a living organism there is an equilibrium. The amount which is absorbed equals the amount which decays. The level of the equilibrium is characteristic for the organism, or a part thereof (e.g. wood from an oak or a human bone). After death no more carbon-14 is absorbed, and the carbon-14 which was present at the time of death decays. The half-life of carbon-14 has been determined to be about 5568 years. For many organisms one also knows how many carbon-14 decays to expect at the time of death. Measuring the number of decays in a dead organism allows us to determine the time of death, approximately. We explain the process in a numerical example.

Example 1.65. Suppose we measure 6.68 carbon-14 decays per minute and gram in a certain kind of wood at the time of death of the tree. Suppose dead wood of the same kind shows 1.8 decays per minute and gram. How long ago did the tree die?

Solution: Let $t_0 = 0$ be the time of death of the tree, and t_1 the present time, measured in years. The number of decays to be expected t years after death is

$$A(t) = 6.68 e^{-\frac{\ln 2}{5568} t}.$$

We have that $A(t_1) = 1.8$. From this we calculate:

$$\ln \frac{1.8}{6.68} = -\frac{\ln 2}{5568} t_1 \quad \text{or} \quad t_1 = -\frac{5568}{\ln 2} \ln\left(\frac{1.8}{6.68}\right) \approx 10,534. \tag{1.49}$$

The tree died approximately $10,500$ years ago.

1.16 More Exponential Growth and Decay

More generally than in (1.45), consider the differential equation

$$f'(t) = a f(t) + b, \tag{1.50}$$

where a and b are constants, and $a \neq 0$. A time independent solution (*steady state solution*) of this equation is $f(t) = -b/a$.

Theorem 1.66. *Functions of the form*

$$f(t) = ce^{at} - \frac{b}{a}$$

are solutions of the differential equation in (1.50)*. Here c denotes an arbitrary constant. On an interval every solution of* (1.50) *is of this form.*

We obtain a unique solution if we add an initial condition to the differential equation in (1.50).

Theorem 1.67. *On an interval which contains t_o, the function*

$$f(t) = \left(y_0 + \frac{b}{a}\right) e^{a(t-t_0)} - \frac{b}{a}$$

is the unique solution of the initial value problem

$$f'(t) = af(t) + b \quad \text{and} \quad f(t_0) = y_0.$$

Remark 2. It is not hard to verify that the given functions are solutions of the respective problems. The uniqueness assertion is a minor modification of Proposition 2.9 on page 68.

Let us apply these ideas to solve some problems. The important aspects are to translate the given information into a mathematical equation. The rest will be routine calculation.

Example 1.68. On graduation day the balance of your student loan is \$15,000. Interest is added at a rate of .5% per month, and you are repaying the loan at a rate of \$ 200.00 per month. Analyze the future of the loan.

Solution: As variable we use time, denoted by t and measured in months. We set $t = 0$ at the time of graduation. This is the time at which you start to repay the loan. Denote the balance of your loan at time t by $B(t)$. The balance increases at a rate of $.005B(t)$ due to interest being added and decreases at a rate of \$200.00 per month due to payments which you make. In summary, we have the initial value problem

$$B'(t) = .005B(t) - 200 \quad \text{and} \quad B(0) = 15,000.$$

According to Theorem 1.67 the solution of the initial value problem is

$$B(t) = \left(15,000 + \frac{-200}{.005}\right) e^{.005t} - \frac{-200}{.005} = -25,000e^{.005t} + 40,000.$$

For example, $B(T) = 0$ if

$$T = \frac{1}{.005} \ln\left(\frac{40}{25}\right) \approx 94.$$

After approximately 94 months (7 years and 10 months) you repaid the loan. Your total payments were \$18,800, so that you paid the principal plus \$3,800 in interest. ◊

Example 1.69. You are absorbing a medication at a rate of 3 mg per hour. (You can keep this rate constant with a skin patch.) The liver metabolizes the medication at a rate of 4% per hour. Analyze the amount of medication in your body at any time.

Solution: We use time as independent variable, denote it by t and measure it in hours. We denote by $t = 0$ the time when we start taking the medication. Let $A(t)$ denote the amount of medication in your body, measured in milligrams. Then $A(t)$ increases at a rate of 3 mg per hour because you are taking in medication and at the same time $A(t)$ decreases at a rate of $.04A(t)$ due to your liver metabolizing the medication. We have the initial value problem

$$A'(t) = -.04A(t) + 3 \quad \text{and} \quad A(0) = 0.$$

The solution of this problem is

$$A(t) = -75e^{-.04t} + 75.$$

For example, after 12 hours there will be about 28.6 milligram of medication in your body. It will take slightly more than 40 hours before the amount of medication in your body reaches 60 milligrams. The steady state solution of the problem is $A(t) = 75$. The amount of medication will stabilize at this amount with time. ◊

Example 1.70 (Newton's Law of Cooling). Suppose you have an object whose temperature is different from the temperature of its surroundings. With time, the temperature of the object will approach the one of its surroundings. We discuss how this happens, at least under idealized circumstances.

Think of a cup of coffee. You stir the coffee gently so that the temperature in the cup remains homogeneous and almost no energy is added through the process of stirring.[10] Denote the temperature of the coffee by T. It is a function of time, so that we write $T(t)$. *Newton's law of cooling* says that the rate at which the heat is transferred, and with this the rate of change of temperature of the coffee, is proportional to the temperature difference. If K is the temperature of the surroundings, then

$$T'(t) = a(T(t) - K) = aT(t) - aK. \tag{1.51}$$

Setting $b = -aK$, this is the differential equation in (1.50).

Let us work out a numerical example. At time $t = 0$, just after you poured the coffee into your cup, its temperature is 95 degrees Celsius. Five minutes later the temperature has dropped to 80 degree, while you stir it slightly and patiently. The room temperature is 25 degrees.

1. Determine the function $T(t)$.

2. Find t_1, such that $T(t_1) = 70$ degrees Celsius.

Solution: To apply Theorem 1.67, we set $t_0 = 0$, $y_0 = 95$, and $K = 25$. Note that $-b/a = K$. Putting all of this into the formula for the solution of the initial value problem, we get that

$$T(t) = (95 - 25)e^{at} + 25 = 70e^{at} + 25.$$

To determine a we use that

$$T(5) = 80 = 70e^{5a} + 25,$$

and we conclude that $a = \frac{1}{5}\ln\left(\frac{55}{70}\right) \approx -.0482$. Using these data, Equation (1.51) says that the temperature of the coffee drops at a rate of about .048 degrees per minute for each degree of difference between the temperature of the coffee and the room temperature. Having a numerical value for a gives us an explicit expression for the temperature T as a function of t:

$$T(t) = 70e^{-.0482t} + 25.$$

[10]The physics of heat transfer changes substantially if you take a solid object, such as a turkey in the oven. The temperature in the solid object will not be homogenous, the outside warms up much faster than the inside. In addition, the specific heat (the amount of energy needed to increase the temperature of one unit of the material by one degree) varies. It is different for fat, protein, and bone. Furthermore, the specific heat is highly temperature dependent for substances like protein. That means, a in (1.51) depends on the temperature T. All of this leads to a significantly different development of the temperature inside a turkey as you roast it for your Thanksgiving dinner.

We like to find out the time t_1 for which

$$T(t_1) = 70e^{-.0482t_1} + 25 = 70.$$

Solving the equation for t_1, we find that $t_1 \approx 9.17$. That means that the temperature drops to 70 degrees approximately 9.17 minutes after pouring it. $\diamondsuit$

Exercise 8. A chemical factory is located on the banks of a river. Down stream from the factory is a lake, and the river is the only contributor to the lake. Assume that the amount of water carried by the river is the same all year around, and the amount of water in the lake is 10 times the amount of water carried by the river per year. In negotiations which the EPA, the owner has agreed to an acceptable level of 2.5 mg per m^3 of a pollutant in the lake. After a major accident the level has risen to 15 mg per m^3. As a remedy, the factory owner proposes to reduce the emission of pollution so that the level of pollutant in the river is only 1.5 mg per m^3. It is assumed that the pollutant is distributed uniformly in the lake at any time.

1. Let $P(t)$ denote the amount of pollutant (measured in mg per m^3) in the lake at time t. Let $t_0 = 0$ be the time just after the accident and at which the clean-up strategy is implemented. State the initial value problem for $P(t)$.

2. Find the function $P(t)$.

3. At which time will the level of pollution be back to 2.5 mg per m^3?

Exercise 9. The population of an endangered species of birds on Kauai decreases at a relative rate of 25% per year. Currently, at time $t_0 = 0$, the population is estimated to be 700 birds. A government agency raises the species in captivity and releases birds into the wild at a rate of 80 birds per year. Denote the size of the population at time t by $P(t)$, where t denotes time and is measured in years.

1. State the initial value problem for $P(t)$.

2. Find the function $P(t)$.

3. At which time will the population drop to 500 birds?

4. What is the long term estimate for the population of this species in the wild?

1.17 The Second and Higher Derivatives

Let $f(x)$ be a function which is defined on an open set. If the function is differentiable at each point of its domain, then $f'(x)$ is again a function with the same domain as $f(x)$. We may ask whether the function $f'(x)$ is differentiable. Its derivative, wherever it exists, is called the second derivative of f. It is denoted by $f''(x)$. This process can be iterated. The derivative of the second derivative is called the third derivative, and denoted by $f'''(x)$, etc. We will make use of the second derivative. Leibnitz's notation for the second derivative of a function $f(x)$ is d^2f/dx^2. Here is a sample computation in which you are invited to fill in the details:

$$\frac{d^2}{dx^2}e^{\sin x} = \frac{d}{dx}\cos x e^{\sin x} = (-\sin x + \cos^2 x)e^{\sin x}.$$

Exercise 10. Find the second derivatives of the following functions:

(1) $f(x) = 3x^3 + 5x^2$	(6) $k(x) = \cos(x^2)$	(11) $p(x) = \ln^2(x + 4)$
(2) $g(x) = \sin 5x$	(7) $l(x) = \ln 2x$	(12) $q(x) = e^{\cos x}$
(3) $h(x) = \sqrt{x^2 + 2}$	(8) $m(x) = \ln(x^2 + 3)$	(13) $r(x) = \ln(\tan x)$
(4) $i(x) = e^{5x}$	(9) $n(x) = \arctan 3x$	(14) $s(x) = e^{x^2-1}$
(5) $j(x) = \tan x$	(10) $o(x) = \sec(x^3)$	(15) $t(x) = \sin^3 x$.

1.18 Numerical Methods

In this section we introduce some methods for numerical computations. Their common feature is, that for a differentiable function we do not make a large error when we use the tangent line to the graph instead of the graph itself. This rather casual statement will become clearer when you look at the individual methods.

1.18.1 Approximation by Differentials

Suppose x_0 is an interior point of the domain of a function $f(x)$ and $f(x)$ is differentiable at x_0. Assume also that $f(x_0)$ and $f'(x_0)$ are known. The method of *approximation by differentials* provides an approximate values $f(x_1)$ if x_1 is near x_0. We use the symbol '≈' to stand for 'is approximately'. One uses the formula

$$f(x_1) \approx f(x_0) + f'(x_0)(x_1 - x_0). \tag{1.52}$$

On the right hand side in (1.52) we have $l(x_1)$, the tangent line to the graph of $f(x)$ at $(x_0, f(x_0))$ evaluated at x_1. In the sense of the definition of the tangent line in Section 1.6, $f(x_1)$ is close to $l(x_1)$ for x_1 near x_0.

Example 1.71. Find an approximate value for $\sqrt[3]{9}$.

Solution: We set $f(x) = \sqrt[3]{x}$, so we are supposed to find $f(9)$. Note that

$$f'(x) = \frac{1}{3}x^{-2/3}, \quad f(8) = 2, \text{ and } f'(8) = \frac{1}{12}.$$

Formula (1.52), applied with $x_1 = 9$ and $x_0 = 8$, says that

$$\sqrt[3]{9} = f(9) \approx 2 + \frac{1}{12}(9-8) = \frac{25}{12} \approx 2.0833.$$

Your calculator will give you $\sqrt[3]{9} \approx 2.0801$. $\diamondsuit$

Example 1.72. Find an approximate value for $\tan 46°$.

Solution: We carry out the calculation in radial measure. Note that $46° = 45° + 1°$, and this corresponds to $\pi/4 + \pi/180$. Use the function $f(x) = \tan x$. Then $f'(x) = \sec^2 x$, $f(\pi/4) = 1$, and $f'(\pi/4) = 2$. Formula (1.52), applied with $x_1 = (\pi/4 + \pi/180)$ and $x_0 = \pi/4$ says

$$\tan 46° = \tan\left(\frac{\pi}{4} + \frac{\pi}{180}\right) \approx \tan\left(\frac{\pi}{4}\right) + \sec^2\left(\frac{\pi}{4}\right)\left(\frac{\pi}{180}\right) = 1 + \frac{\pi}{90} \approx 1.0349.$$

Your calculator will give you $\tan 46° \approx 1.0355$. $\diamondsuit$

Exercise 11. Use approximation by differentials to find approximate values for

$$(1)\ \sqrt[5]{34} \quad (2)\ \tan 31° \quad (3)\ \ln 1.2 \quad (4)\ \arctan 1.1.$$

In each case, compare your answer with one found on your calculator.

We have been causal in (1.52) insofar as we have not estimated (provided an upper bound for) the *error* which we make using the right hand side of (1.52) instead of of the actual value of the function on the left hand side. The inequality in Definition 1.25 on page 12 provides us with an estimate. According to this slightly more demanding definition, differentiability of the function $f(x)$ means that there exist numbers A and $d > 0$, such that

$$|f(x_1) - [f(x_0) + f'(x_0)(x_1 - x_0)| \le A(x_1 - x_0)^2$$

whenever $|x_1 - x_0| < d$. Thus, if we know A and d, then we can approximate the error as long as $|x_1 - x_0| < d$.

Example 1.73. Find an approximate value for $\sin 31°$ and estimate the error.

Solution: Set $f(x) = \sin x$. Then $f'(x) = \cos x$, $f(\pi/6) = 1/2$, and $f'(\pi/6) = \sqrt{3}/2$. Measuring angles in radians we set $x_0 = \pi/6$ and $x_1 = \pi/6 + \pi/180$. Applying the formula in (1.52), we find

$$\sin 31° \approx \sin\frac{\pi}{6} + \frac{\pi}{180}\cos\frac{\pi}{6} = \frac{1}{2}\left(1 + \sqrt{3}\frac{\pi}{180}\right) \approx .515115.$$

The calculator will tell that $\sin 31° \approx .515038$.

From the computation in Example 1.31 on page 17 we also know that we may use $A = 1$ and $d = \pi/4$ in the differentiability estimate. The estimate assures us that the error is at most

$$(x_1 - x_0)^2 = \left(\frac{\pi}{180}\right)^2 \leq .000305.$$

Comparison of the actual and approximate value confirm this. $\diamondsuit$

Example 1.74. Use approximation by differentials to find an approximate value of $\sqrt{10}$ and give an upper bound for the error.

Solution: We use $f(x) = \sqrt{x}$ and $x_0 = 9$. The $f'(x) = 1/(2\sqrt{x})$, $f(x_0) = 3$, and $f'(x_0) = 1/6$. The formula in (1.52) tells us that

$$\sqrt{10} = f(10) \approx f(9) + f'(9)(10-9) = 3 + \frac{1}{6} \approx 3.16666.$$

The calculator will give you $\sqrt{10} \approx 3.16228$.

For the error estimate we may use

$$A = \frac{1}{2(\sqrt{x_0})^3}$$

and any $d > 0$, see (1.16). The estimate assures us that the error is at most

$$\frac{1}{2(\sqrt{x_0})^3}(x_1 - x_0)^2 = \frac{1}{54}.$$

The actual error is again substantially less than this. $\diamondsuit$

Exercise 12. Use approximation by differentials to find approximate values for

$$(1)\ \cos 28° \quad (2)\ \sqrt{26} \quad (3)\ \sin 47°.$$

In each case, estimate also the maximal error which you may have made by using the method of approximation by differentials.

1.18.2 Newton's Method

Newton's method is designed to find the zeros of a function. You have learned how to solve linear and quadratic equations, i.e., finding the zeros of functions of degree 1 and 2. More sophisticated methods allow you to find the exact solutions of polynomial equations of degree three and four. For polynomials of degree greater or equal to 5 and most other functions there are no general methods for finding their roots.

Newton's method works as follows. Suppose we want to find a zero of a differentiable function $f(x)$, i.e., we want to find some $\overline{x}$, such that $f(\overline{x}) = 0$. Suppose that by some means we know that such an $\overline{x}$ exists, and that x_0 is not far from $\overline{x}$. Then we set

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}, \quad x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}, \text{ etc.} \tag{1.53}$$

and in general

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{1.54}$$

Geometry of Newton's Method: Let us give a geometric explanation for the formlas. Given any x_0 at which f is defined and differentiable, we obtain the tangent line $l(x)$ to the graph of f at this point. Then x_1, as given in (1.53), is the point at which $l(x)$ intersects the x-axis. Specifically,

$$l(x) = f'(x_0)(x - x_0) + f(x_0), \quad \text{and} \quad l(x_1) = 0 \quad \text{if} \quad x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

This means that we accept that the tangent line is close to the graph of the function, and instead of finding the zero of the function itself, we find the zero of the tangent line. The process is then iterated.

Let us calculate $\sqrt{A}$, i.e., the positive root of the function $f(x) = x^2 - A$. Then $f'(x) = 2x$, and

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^2 - A}{2x_n} = \frac{x_n^2 + A}{2x_n} = \frac{1}{2}\left(x_n + \frac{A}{x_n}\right) \tag{1.55}$$

If we use $A = 3$ and $x_0 = 2$, then we find

$$x_1 = \frac{1}{2}\left(2 + \frac{3}{2}\right) = \frac{7}{4}, \quad x_2 = \frac{1}{2}\left(\frac{7}{4} + \frac{12}{7}\right) = \frac{97}{56} \quad \text{and} \quad x_3 = \frac{18817}{10864}.$$

We summarize the computation in Table 1.1. In the first column you find the subscript n. In the following two columns you find the values of x_n, once expressed as a fraction of integers, once in decimal form. In the last column you see the square of x_n. At least x_3^2 is rather close to 3. Your calculator will give you 1.73205080757 as an approximate value of $\sqrt{3}$. You see that our value for x_3 is rather precise. In fact, if you carry the calculation one step further and find x_4, then the accuracy of this approximation of $\sqrt{3}$ will exceed the accuracy of most calculators. The numbers in the last column show that we are making rapid progress in finding a good approximation of $\sqrt{3}$.

n	x_n	x_n	x_n^2
0	2	2.0000000000	4.0000000000
1	7/4	1.7500000000	3.0625000000
2	97/56	1.7321428571	3.0003188775
3	18817/10864	1.7320508100	3.0000000085

Table 1.1: The Babylonian Method

More than 4000 years ago the Babylonians used the outermost expressions in (1.55)

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{A}{x_n}\right)$$

to find good approximations of square roots, expressed as rational numbers. We refer to the described procedure as the *Babylonian method*.

Let us consider one more example to illustrate Newton's method. Find a solution of the equation

$$x \sin x = \cos x.$$

Equivalently, we may say, find a root of the function $f(x) = x \sin x - \cos x$.

Step 1: Let us make sure that there is a root of the function to be found. Observe that $f(0) = -1 < 0$ and $f(\pi/2) = \pi/2 > 0$. The intermediate value theorem tells us that $f(x)$, as a continuous funtion, has a root in the interval $(0, \pi/2)$. Let us call this root $\overline{x}$.

Step 2: Let us come up with a first guess for a root. Considering the values of f at the end points of the interval, we guess that $x_0 = 1$ is not too far away from the root, which we know to exist by Step 1. Actually $f(1) \approx .3$.

Step 3: Let us improve the guess: Set

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \approx .8645.$$

Your calculator will tell you that $f(x_1) \approx .00874$. You see that $f(x_1)$ is much closer to zero than $f(x_0)$, and in this sense we expect that x_1 is much closer to the root $\overline{x}$ of $f(x)$ than x_0. We made progress finding $\overline{x}$.

Step 4: Repeat Step 3 and calculate x_2, x_3, ... The distance between $\overline{x}$ and x_n will decrease rapidly as n increases.

We explained Newton's method because we want to illustrate the power of the concept '*tangent line*.' A full discussion of Newton's method requires mathematical tools which are not available to us at this time. In general, many interesting phenomena can occur. Still, the principle problem is as follows. Suppose that $f(x)$ is a differentiable function and $f(\overline{x}) = 0$ and x_0 is given. Suppose that x_n for $n \geq 1$ are computed according to (1.54). Do the x_n tend (converge) to $\overline{x}$, and how fast?

For completeness sake, we give an answer. Consider an interval $I = [\overline{x} - a, \overline{x} + a]$, and suppose that $|f'(x)| \geq m$ and $|f''(x)| \leq M$ on I. Suppose $x_n \in I$ and $|\overline{x} - x_n| < \frac{2m}{M}$. A theorem from advanced calculus asserts that

$$|\overline{x} - x_{n+1}| \leq \frac{M}{2m}(\overline{x} - x_n)^2. \tag{1.56}$$

We illustrate the theorem by applying it to the previous example. Observe that $f(.8) < 0$ and $f(.9) > 0$. This tells us that $\overline{x} \in [.8, .9]$. Let us set $a = .2$, so that $I \subset J = [.6, 1.1]$. On J, and with this also on I, we have that $|f'(x)| \geq m = 1.5$ and $|f''(x)| \leq M = 2.5$. You are invited to verify these estimates using technology. In (1.56) we use that $\frac{M}{2m} < 1$. As first guess we used $x_0 = 1$, so that we know that $|\overline{x} - x_0| < .2$. The quoted theorem asserts that $|\overline{x} - x_1| < .04$. If we repeat the process, then we see that $|\overline{x} - x_2| < .0016$ and $|\overline{x} - x_3| < .00000256$. This illustrates that the x_n approach $\overline{x}$ rapidly.

There is one feature of Newton's method which helps. You may say that with each iteration you make a fresh start, and in this sense previous round-off errors don't carry over.

1.18.3 Euler's Method

Euler's method is designed to find, by numerical means, an approximate solution of the following kind of problem:

Problem 1. *Find a function $y(t)$ which satisfies*

$$y' = \frac{dy}{dt} = F(t, y) \quad \textit{and} \quad y(t_0) = y_0. \tag{1.57}$$

Here $F(t, y)$ denotes a given function in two variables, and t_0 and y_0 are given numbers.

The first condition on y in (1.57) is a *first order differential equation*. It is an equation which involves a function and its derivative, and the unknown is the function. The second condition is called an *initial condition*. It specifies the value of the function at one point. For short, the problem in (1.57) is called an *initial value problem*.

Approach in one step: Suppose you want to find $y(T)$ for some $T \neq t_0$. Then you might try the formula

$$y(T) \approx y(t_0) + y'(t_0)(T - t_0) = y_0 + F(t_0, y_0)(T - t_0). \tag{1.58}$$

The tangent line to the graph of y at (t_0, y_0) is

$$l(t) = y(t_0) + y'(t_0)(t - t_0),$$

so that the middle term in (1.58) is just $l(T)$. The first, approximate equality in (1.58) expresses the philosophy that the graph of a differentiable function is close to its tangent line, at least as long as T is close to t_0. To get the second equality in (1.58) we use the differential equation and initial condition in (1.57), which tell us that

$$y'(t_0) = F(t_0, y(t_0)) = F(t_0, y_0).$$

The Logistic Law

The differential equation in our next example is known as the *logistic law of population growth*. In the equation, t denotes time and $y(t)$ the size of a population, which depends on t. The constants a and b are called the vital coefficients of the population. The equation was first used in population studies by the Dutch mathematician-biologist Verhulst in 1837. The equation refines the Malthusian law for population growth (see (1.45)).

In the differential equation, the term ay expresses that population growth is proportional to the size of the population. In addition, the members of the population meet and compete for food and living space. The probability of this happening is proportional to y^2, so that it is assumed that population growth is reduced by a term which is proportional to y^2.

Example 1.75. Consider the initial value problem:

$$\frac{dy}{dt} = ay - by^2 \quad \text{and} \quad y(t_0) = y_0, \tag{1.59}$$

where a and b are given constants. Find an approximate value for $y(T)$.

Remark 3. An exact solution of the initial value problem in (1.59) is given by the equation

$$y(t) = \frac{ay_0}{by_0 + (a - by_0)e^{-a(t-t_0)}} \tag{1.60}$$

This is not the time to derive this exact solution, though you are invited to verify that it satisfies (1.59). We are providing the exact solution, so that we can see how well our approximate values match it.

Solution: Setting $F(t, y) = ay - by^2$, you see that the differential equation in this example is a special case of the one in (1.57). According to the formula in (1.58) we find

$$y(T) \approx y_0 + (ay_0 - by_0^2)(T - t_0). \tag{1.61}$$

We expect a close approximation only for T close to t_0. ♢

Let us be even more specific and give a numerical example.

Example 1.76. Consider the initial value problem.

$$\frac{dy}{dt} = \frac{1}{10}y - \frac{1}{10000}y^2 \quad \text{and} \quad y(0) = 300. \tag{1.62}$$

Find approximate values for $y(1)$ and $y(10)$.

Solution: Substituting $a = 1/10$, $b = 1/10000$, $t_0 = 0$, and $y_0 = 300$ into the solution in (1.61), we find that

$$y(1) \approx 300 + \left(\frac{300}{10} - \frac{300^2}{10000}\right)(1 - 0) = 321.$$

According to the exact solution in (1.60), we find that

$$y(t) = \frac{3000}{3 + 7e^{-t/10}}.$$

Substituting $t = 1$, we find the exact value $y(1) = 321.4$; this number is rounded off. So, our approximate value is close.

For $T = 10$ the formula suggests that $y(10) \approx 510$. According to the exact solution for this initial value problem, $y(10) = 538.1$. For this larger value of T, the formula in (1.61) gives us a less satisfactory result. ♢

Multi-step approach: We like to find a remedy for the problem which we discovered in Example 1.76 for T further away from t_0. Consider again Problem 1 on page 54. We want to get an approximate value for $y(T)$. For notational convenience we assume that $T > t_0$. Pick several t_i between t_0 and T:

$$t_0 < t_1 < t_2 < \cdots < t_n = T.$$

Starting out with t_0 and $y(t_0)$, we use the one step method from above to get an approximate value for $y(t_1)$. Then we pretend that $y(t_1)$ is exact, and we repeat the process. We use t_1 and $y(t_1)$ to calculate an approximate value for $y(t_2)$. Again we pretend that $y(t_2)$ is exact and use t_2 and $y(t_2)$ to calculate $y(t_3)$. Iteratively, we calculate $[t_{i+1}, y(t_{i+1})]$ from $[t_i, y(t_i)]$ according to the formula in (1.58):

$$[t_{i+1}, y(t_{i+1})] = [t_{i+1}, y(t_i) + F(t_i, y(t_i))(t_{i+1} - t_i)] \tag{1.63}$$

We continue this process until we reach T.

For reasonably nice[11] expressions $F(t, y)$ the accuracy of the value which we get for $y(T)$ will increase with n, the number of steps we make (at least if all steps are of the same length). On the other hand, in an actual numerical computation we also make round-off errors in each step, and the more steps we make the worse the result might get. Experience will guide you in the choice of the step length.

Example 1.77. Consider the initial value problem

$$\frac{dy}{dt} = \frac{1}{10}y - \frac{1}{10000}y^2 \quad \text{and} \quad y(0) = 10. \tag{1.64}$$

1. Apply the multi-step method to find approximate values for $y(t)$ at $t_1 = 5$, $t_2 = 10$, $t_3 = 15$, ... , $t_{20} = 100$. Arrange them in a table.

2. Graph the points found in the previous step together with the actual solution of the initial value problem.

Solution: As points in the multi-step process we use

$$t_0 = 0,\ t_1 = 5,\ t_2 = 10,\ t_3 = 15,\ t_4 = 20,\ \ldots,\ t_{20} = 100.$$

[11] We do not want to make this term precise, but the $F(t, y)$ in Example 1.75 is of this kind.

t	$y(t)$	&	t	$y(t)$	&	t	$y(t)$
0	10.00		35	153.96		70	857.73
5	14.95		40	219.09		75	918.74
10	22.31		45	304.62		80	956.07
15	33.22		50	410.55		85	977.07
20	49.28		55	531.55		90	988.27
25	72.70		60	656.05		95	994.07
30	106.41		65	768.87		100	997.02

Table 1.2: Solution of Problem 1.77

For each t_i $(0 \leq i \leq 19)$ we use the formula

$$y(t_{i+1}) = y(t_i) + 5\left(\frac{y(t_i)}{10} - \frac{y_i^2(t_i)}{10000}\right)$$

and calculate $y(t_1)$, $y(t_2)$, $y(t_3)$, $\ldots$, $y(t_{20})$ consecutively. We summarize the calculation in Table 1.2.

In Figure 1.15 you see the graph of the exact solution of the initial value problem. You also see the points from Table 1.2. The points suggest a graph which does follow the actual one reasonably closely. But you see that we are definitely making errors, and they get worse as t increases[12]. You may try a shorter step length. The points will follow the curve much more closely if you use $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $\ldots$, $t_{100} = 100$ in your calculation. $\Diamond$

Steady States: Let us consider some very specific solutions of our initial value problem in (1.57):

$$y' = \frac{dy}{dt} = F(t, y) \quad \text{and} \quad y(t_0) = y_0.$$

Suppose $F(y_0, t) = 0$ for all t. Then the constant function $y(t) = y_0$ is a solution of the problem. Such a solution is called a *steady state solution.*

[12] It is incidental that the points eventually get closer to the graph again. This is due to the specific problem, and will not occur in general.

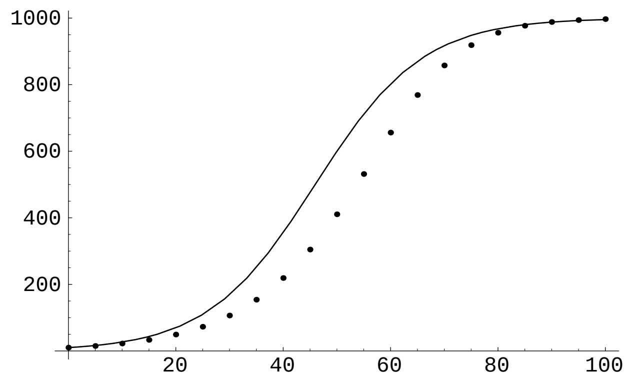

Figure 1.15: Illustration of Euler's Method

Example 1.78. Find the steady states of the differential equation (see (1.50) in Section 1.16)

$$f'(t) = af(t) + b. \tag{1.65}$$

Solution: Apparently, $f'(t) = 0$ if and only if $f(t) = -b/a$. So the constant function $f(t) = -b/a$ is the only steady state of this differential equation. $\diamondsuit$

In review of Example 1.68 in Section 1.16, you see that the steady state in that example is $B(t) = 40,000$. I.e., if your loan balance is \$40,000.00, the bank charges you interest at a rate of .5% per month, and you are repaying the loan at a rate of \$ 200.00 per month, then the principal balance of your account will stay unchanged. Your payments cover exactly the occuring interest charges.

Example 1.79. For the logistic law (see Equation (1.59))

$$\frac{dy}{dt} = F(y,t) = ay - by^2 = y(a - by)$$

we find that $F(y,t) = 0$ if and only if $y = 0$ or $y = a/b$. There are two steady state solutions: $y_u(t) = 0$ and $y_s(t) = a/b$.

Let us interpret these steady state solutions for the specific numerical values of $a = 1/10$ and $b = 1/10,000$ in Example 1.77. If the initial value y_0 of the population is positive, then the population size will tend to and stabilize[13] at $y(t) = a/b = 1,000$. In this sense, $y_s(t) = a/b = 1,000$ is a stable steady state solution. It is also referred to as the *carrying capacity.* It tells you which size population of the given kind the specific habitat will support.

If the initial value y_0 is negative, then $y(t)$ will tend to $-\infty$ as time increases. If $y_0 \neq 0$, then $y(t)$ will not tend to the steady state $y(t) = 0$. In this sense, $y(t) = 0$ is an unstable steady state. $\Diamond$

Exercise 13. Consider the initial value problem

$$y'(t) = -50 + \frac{1}{2}y(t) - \frac{1}{2000}y^2(t) \quad \text{and} \quad y_0 = y(0) = 200. \tag{1.66}$$

To make the problem explicit, you should think of a population of deer in a protected wildlife preserve. There are no predators. The deer are hunted at a rate of 50 animals per year. The population has a growth rate of 50% per year. Reproduction takes place at a constant rate all year round. Finally, the last term in the differential equation accounts for the competition for space and food.

1. Use Euler's method to find the population size over the next 30 years. Proceed in 1 year steps. Tabulate and plot your results.

2. Guess at which level the population stabilizes.

3. Repeat the first two steps of the problem if hunting is stopped.

4. Repeat the first two steps of the problem if the initial population is 100 animals.

5. Find the steady states of the original equation in which hunting takes place. I.e., find for which values of y you have that $y' = 0$? You will find two values. Call the smaller one of them Y_u and the larger one Y_s. Experiment with different initial values to see which of the steady states is stable, and which one is unstable.

[13] The common language meaning of these expressions suffices for the purpose of our discussion, and the mathematical definition of *'tends to'* and *'stabilizes at'* only make these terms precise.

Orthogonal Trajectories

Let us explore a different kind of application. Suppose we are given a family $F(x, y, a) = 0$ of curves. In Figure 1.16 you see a family of ellipses

$$C_a : \ F(x, y, a) = x^2 + 3y^2 - a = 0. \tag{1.67}$$

There is one ellipse for each $a > 0$. We like to find curves D_b which intersect the curves C_a perpendicularly. (We say that D_b and C_a intersect perpendicularly in a point (x_1, y_1), if the tangent lines to the curves at this point intersect perpendicularly.) We call such a curve D_b an *orthogonal trajectory* to the family of the C_a's. You also see one orthogonal trajectory in Figure 1.16.

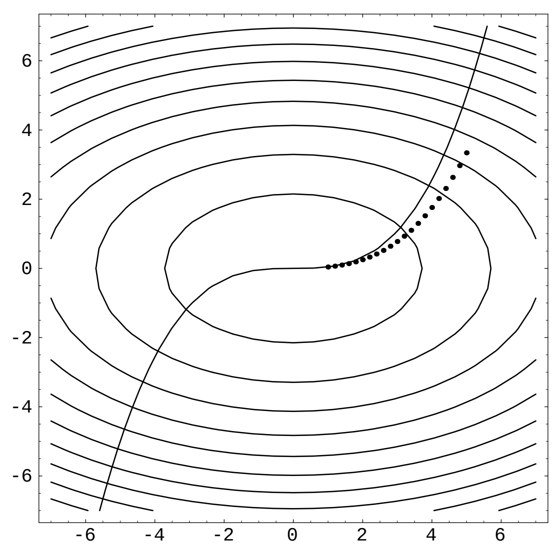

Figure 1.16: Orthogonal Trajectory to Level Curves

Let us explain where this type of situation occurs. Suppose the curves C_a are the level curves in a crater. Here a represents the elevation, so that the elevation is constant along each curve C_a. The orthogonal trajectory gives a path of steepest descent. A new lava flow which originates at some point in the crater will follow this path.

Suppose that each ellipse represents an equipotential line of an electromagnetic field. The orthogonal trajectory provides you with a path which is always in the direction of the most rapid change of the field. A charged particle will move along an orthogonal trajectory.

Suppose a stands for temperature, so that along each ellipse the temperature is constant. In this case the curves are called isothermal lines[14]. A heat seeking bug will, at any time, move in the direction in which the temperature increases most rapidly, i.e., along an orthogonal trajectory to the isothermal lines.

Suppose a stands for the concentration of a nutrient in a solution. It is constant along each curve C_a. On their search for food, bacteria will follow a path in the direction in which the concentration increases most rapidly. They will move along an orthogonal trajectory.

Example 1.80. Find orthogonal trajectories for the family of ellipses

$$C_a : \ F(x, y, a) = x^2 + 3y^2 - a = 0. \tag{1.68}$$

Solution: Differentiating the equation for the ellipses, we get

$$2x + 6y\frac{dy}{dx} = 0 \quad \text{or} \quad \frac{dy}{dx} = \frac{-x}{3y}.$$

The slope of the tangent line to a curve C_a at a point (x_1, y_1) is $\frac{-x_1}{3y_1}$. If a curve D_b intersects C_a in (x_1, y_1) perpendicularly, then we need that the slope of the tangent line to D_b at this point is $\frac{3y_1}{x_1}$. Thus, to find an orthogonal trajectory to the family of the C_a's we need to find functions which satisfy this differential equation. If we also require that the orthogonal trajectory goes through a specific point (x_0, y_0), then we end up with the initial value problem

$$\frac{dy}{dx} = \frac{3y}{x} \quad \text{and} \quad y(x_0) = y_0.$$

This is exactly the kind of problem which we solved with Euler's method. In this particular example it is not difficult to find solutions for the differential equation. They are functions of the form $y(x) = bx^3$. The orthogonal trajectory shown in Figure 1.16 has the equation $y = x^3/25$. There is one orthogonal trajectory which does not have this form, and this is the curve $x = 0$.

Let us apply Euler's method to solve the problem. Let us find approximate values for the initial value problem

$$\frac{dy}{dx} = \frac{3y}{x} \quad \text{and} \quad y(1) = \frac{1}{25}.$$

[14]The idea of isothermal lines, and with this the method in all of these applications, was pioneered by Alexander von Humbold (1769–1859).

Use $x_0 = 1$, $x_1 = 1.2$, $x_2 = 1.4$, ... , $x_{20} = 5$.

We set $(x_0, y_0) = (1, 1/25)$ and calculate (x_n, y_n) according to the formula

$$y_n = y_{n-1} + .2\frac{3y_{n-1}}{x_{n-1}} \quad \text{for } n = 1, 2, \ldots, 20.$$

Without recording the results of this calculation, we graphed the points in Figure 1.16. ♢

Exercise 14. Consider the family of hyperbolas:

$$C_a : \ x^2 - 5y^2 + a = 0.$$

There is one hyperbola for each value of a, only for $a = 0$ the hyperbola degenerates into two intersecting lines.

1. Graph several of the curves C_a.
2. Find the differential equation for an orthogonal trajectory.
3. Use Euler's method to find points on the orthogonal trajectory through the point $(3, 4)$. Use the points $x_0 = 3$, $x_1 = 3.2$, $x_2 = 3.4$, ... , $x_{20} = 7$. Plot the points (x_n, y_n) in your figure.
4. Check that the graph of $y(x) = bx^{-5}$ is an orthogonal trajectory to the family of hyperbolas for every b. Determine b, so that the orthogonal trajectory passes through the point $(3, 4)$, and add this graph to your figure.

1.19 Table of Important Derivatives

$f(x)$	$f'(x)$	Assumptions
x^q	qx^{q-1}	q a natural number, or $x > 0$
e^x	e^x	$x \in (-\infty, \infty)$
$\ln \|x\|$	$1/x$	$x \in (-\infty, \infty),\ x \neq 0$
$\sin x$	$\cos x$	$x \in (-\infty, \infty)$
$\cos x$	$-\sin x$	$x \in (-\infty, \infty)$
$\tan x$	$\sec^2 x$	all x for which $\tan x$ is defined
$\cot x$	$-\csc^2 x$	all x for which $\cot x$ is defined
$\sec x$	$\sec x \tan x$	all x for which $\sec x$ is defined
$\csc x$	$-\csc x \cot x$	all x for which $\csc x$ is defined
$\arctan x$	$\frac{1}{1+x^2}$	$x \in (-\infty, \infty)$
$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$	$x \in (-1, 1),\ \arcsin x \in (-\pi/2, \pi/2)$
$\arccos x$	$\frac{-1}{\sqrt{1-x^2}}$	$x \in (-1, 1),\ \arccos x \in (0, \pi)$
$\text{arccot}\, x$	$\frac{-1}{1+x^2}$	$x \in (-\infty, \infty),\ \text{arccot}\, x \in (0, \pi)$
$\text{arcsec}\, x$	$\frac{1}{\|x\|\sqrt{x^2-1}}$	$x < -1$ or $x > 1,\ \text{arcsec}\, x \in (0, \pi/2) \cup (\pi/2, \pi)$
$\text{arccsc}\, x$	$\frac{-1}{\|x\|\sqrt{x^2-1}}$	$x < -1$ or $x > 1,\ \text{arcsec}\, x \in (-\pi/2, 0) \cup (0, \pi/2)$

Table 1.3: Some Derivatives

Chapter 2

Global Theory

So far we studied the local behaviour of a function. All concepts related to the behaviour of a function near a point. In this chapter we will use local information about a function to draw global conclusions. We will discuss some uniqueness properties of solutions of differential equations. Then we discuss geometric properties of graphs, their monotonicity and concavity. We apply these ideas to the study of extrema of functions. With this information it is possible to sketch graphs capturing their essential features.

The fundamental result which allows us to do this is referred to as Cauchy's mean value theorem. Augustin-Louis Cauchy (1789–1857) was one of the great mathematicians of the 19-th century. He made major contributions to make calculus a rigorous mathematical theory.

2.1 Cauchy's Mean Value Theorem

It is useful to make the following

Definition 2.1. *Let $f(x)$ be a function which is defined on the interval $[a, b]$. Then we call*

$$\frac{f(b) - f(a)}{b - a}$$

the average rate of change *of f over the interval $[a, b]$.*

For example, the average rate of change of $f(x) = x^2$ over the interval $[0, 2]$ is 2. The average rate of change of $f(x) = \sin x$ over $[0, \pi/2]$ is $2/\pi$ and over $[0, \pi]$ it is 0.

Theorem 2.2 (Chauchy's Mean Value Theorem). *Let f be a real valued function which is defined and continuous on the interval $[a, b]$ and differentiable on (a, b), where $a < b$. Then there exists a number $c \in (a, b)$ such that*

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

In words, the theorem asserts that the average rate of change over an interval is equal to the rate of change at some point in the interval. For example, the average rate of change of $f(x) = x^2$ over the interval $[-2, 1]$ is -1, and $f'(-1/2) = -1$.

The following special case of the theorem, called Rolle's theorem (named after Michel Rolle (1652–1719)), is of particular interest.

Theorem 2.3 (Rolle's Theorem). *Let f be a real valued function which is defined and continuous on the interval $[a, b]$ and differentiable on (a, b), where $a < b$. If $f(a) = f(b)$, then there exists a number c between a and b (i.e., $a < c < b$) such that*

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

We are not going to say anything about the proof of these two theorems, except that Cauchy's theorem and Rolle's theorem are equivalent (each is an easy consequence of the other one), and that the proof of both of them depends heavily on the completeness[1] of the real numbers. We are also not interested in finding the points c, as they occur in the two theorems. We are interested in more general consequences.

Corollary 2.4. *Let f be a real valued function which is defined and continuous on an interval I. If $f'(x) = 0$ for all interior points x of I, then f is constant on this interval. In other words, there exists a number d such that $f(x) = d$ for all $x \in I$.*

Proof. A different formulation of the claim is that $f(a) = f(b)$ for all a, $b \in I$. We prove this statement using Cauchy's theorem. If $f(a) \neq f(b)$, then $a \neq b$ and there exists some $c \in (a, b)$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} \neq 0.$$

But this contradicts the assumption that $f'(c) = 0$ for all $c \in I$, and the corollary is proved. □

[1] We discussed this property of the real numbers in Section 1.1.

We are going to use the following corollary frequently.

Corollary 2.5. *Let h and g be functions which are defined and continuous on an interval I. If $h'(x) = g'(x)$ for all $x \in I$, then h and g differ by a constant, i.e., there exists a number d such that*

$$h(x) = g(x) + d$$

for all $x \in I$.

Proof. Apply the previous corollary to $f(x) = h(x) - g(x)$. □

Definition 2.6. *Suppose the function $f(x)$ is defined on the interval I. We call a function $F(x)$ with domain I an* antiderivative *of f if $F'(x) = f(x)$ for all $x \in I$.*

Using this notion, we can reformulate Corollary 2.5.

Corollary 2.7. *Suppose h and g are antiderivatives of a function f, defined on an interval. Then h and g differ by a constant.*

2.2 Unique Solutions of Differential Equations

Corollary 2.4 implies

Proposition 2.8. *If the function $F(x)$ is defined on an interval I and $F'(x) = 0$ for all $x \in I$, then $F(x)$ is constant on I.*

In other words, on intervals the only solutions of the differential equation $F'(x) = 0$ are the constant functions.

More generally, if you like to find all antiderivatives $F(x)$ of a function $f(x)$ on an interval, then it suffices to find one antiderivative $H(x)$. Any antiderivative $F(x)$ is of the form $H(x) + c$ where c is a constant. The constant c is referred to as *integration constant*. For the time being you depend on being able to guess such a function $H(x)$. By differentiating $H(x)$ you can check whether you guessed right.

For example, any antiderivative $F(x)$ of the function $f(x) = 2x$ on the real line $(-\infty, \infty)$ is of the form $F(x) = x^2 + c$ where c is a constant. Any antiderivative $F(x)$ of the function $f(x) = \sec^2 x$ on the interval $(-\pi/2, \pi/2)$ is of the form $F(x) = \tan x + c$.

Typically, the integration constant is determined by an initial condition. Suppose we like to solve the initial value problem

$$f'(x) = \cos x \quad \text{and} \quad f(0) = 1.$$

Our first conclusion is that $f(x) = \sin x + c$. This follows from the above because $(\sin x + c)' = \cos x$. Next we substitute $x = 0$ in the equation. Then we see that $f(0) = c = 1$. The solution of the initial value problem is $f(x) = \sin x + 1$.

Of particular importance to our discussion of expenential growth and decay is

Proposition 2.9. *Every solution $f(x)$ of the differential equation*

$$f'(x) = af(x)$$

on an interval is of the form $f(x) = ce^{ax}$ for some constant c.

Proof. We asserted in (1.19), and will eventually prove, that all functions of the form $f(x) = ce^{ax}$ satisfy the differential equation. We want to see that these are the solutions.

Let $f(x)$ be any function which satisfies the differential equation on some interval. Consider the function

$$h(x) = f(x)e^{-ax}.$$

As a product of differentiable functions, h is differentiable. Its derivative is

$$h'(x) = f'(x)e^{-ax} - af(x)e^{-ax} = af(x)e^{-ax} - af(x)e^{-ax} = 0.$$

Corollary 2.4 tells us that $h(x)$ is a constant function. Calling the constant c we find that

$$f(x) = ce^{ax}.$$

This means that all solutions of the differential equation $f'(x) = af(x)$ are of the form $f(x) = ce^{ax}$, where c is a constant. □

Proposition 2.10. *The initial value problem*

$$f'(x) = af(x) \quad \text{and} \quad f(x_0) = C$$

has a unique solution on an interval containing x_0. In fact

$$f(x) = Ce^{a(x-x_0)}.$$

Proof. By the previous proposition we know that the solution is of the form $f(x) = ce^{ax}$ for some c. Substituting the initial condition we obtain

$$C = f(x_0) = ce^{ax_0}.$$

Thus $c = Ce^{-ax_0}$ and $f(x) = ce^{ax} = Ce^{-ax_0}e^{ax} = Ce^{a(x-x_0)}$. □

Remark 4. The uniqueness of the solution of an initial value problem as in the previous proposition is not only of theoretical importance. Imagine that you study the growth rate of a strain of bacteria, as we did in Example 1.64 on page 41. Before you can publish your result, it must be certain that your experiment can be reproduced at a different time in a different location. That is a requirement which any experiment in science must satisfy. If there is more than one mathematical solution to your problem, then you have to expect that the experiment can go either way, and this would invalidate your experiment.

2.3 The First Derivative and Monotonicity

One of the interesting properties of a function is whether it is increasing or decreasing. We might want to find out whether the part of a population which is infected with a disease is increasing or decreasing. We might want to know how the level of pollution in a body of water is changing. The first derivative of a function gives us information of this kind.

2.3.1 Monotonicity on Intervals

Recall that a function f is called *increasing* if $f(b) > f(a)$ whenever $b > a$. It is called *decreasing* if $f(b) < f(a)$ whenever $b > a$. A function is called *monotonic* if it is either increasing or decreasing.

Theorem 2.11. *Suppose that the function f is defined and continuous on the interval I.*

1. *If $f'(x) > 0$ for all $x \in I$, then f is increasing on I.*
2. *If $f'(x) < 0$ for all $x \in I$, then f is decreasing on I.*
3. *More generally, the conclusions in (1) and (2) still hold if in each finite interval $J \subset I$ there are only finitely many points at which the assumption on $f'(x)$ is not satisfied.*[2]

Proof. We show (1). Let a and b be points in I, and suppose that $a < b$. Cauchy's theorem says that there exists a point c, $a < c < b$, such that

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

[2]It is permissable that f is not differentiable at a few points in J, or that $f'(x) = 0$. It is not possible that $f'(x) < 0$ at some point in the interval, and $f(x)$ is increasing on the interval.

We have that $f'(c) > 0$ and $b - a > 0$, and it follows that $f(b) - f(a) > 0$. This means that $f(b) > f(a)$. The proof of the second claim is similar. We leave it and the generalization of both statements to the reader. □

For example, $\log_2' x = \left(\frac{\ln x}{\ln 2}\right)' = \frac{1}{x \ln 2} > 0$ for all $x \in (0, \infty)$. In the computation we used (1.19), Theorem 1.36, and that $\ln 2 > 0$. It follows from Theorem 2.11 that $\log_2 x$ is increasing on $x \in (0, \infty)$. You see part of the graph of the function in Figure 1.8.

The exponential function $\exp_a x = a^x$ is increasing on $(-\infty, \infty)$ if $a > 1$ and decreasing if $0 < a < 1$. To see this, observe that $a^x = e^{x \ln a}$ and $\frac{d}{dx} a^x = (\ln a) a^x$. Furthermore, $a^x > 0$ and $\ln a > 0$ if $a > 1$ and $\ln a < 0$ if $0 < a < 1$. Now Theorem 2.11 implies our assertion. You may also want to have a look at the graph of the exponential function with base 2 in Figure 1.7.

The function $f(x) = 1/x$ is defined and differentiable on the set of all nonzero real numbers, and its derivative is $f'(x) = -1/x^2$. In particular $f'(x) < 0$ for all nonzero real numbers. According to Theorem 2.11, $f(x)$ is decreasing on the interval $(-\infty, 0)$, and that $f(x)$ is decreasing on the interval $(0, \infty)$. The function is not decreasing on the union of the two intervals. The example illustrates that it is crucial in Theorem 2.11 that we deal with functions which are defined and differentiable on an interval.

The function $f(x) = \tan x$, defined on $(-\pi/2, \pi/2)$, has as its derivative $f'(x) = \sec^2 x$, and the derivative is positive. Consequently, $f(x)$ is increasing on $(-\pi/2, \pi/2)$. Its inverse $g(x) = \arctan x$, defined on $(-\infty, \infty)$, has as its derivative $g'(x) = \frac{1}{1+x^2}$, which is positive on $(-\infty, \infty)$, so that $g(x) = \arctan x$ is increasing on $(-\infty, \infty)$. As a general priciple, one may show that the inverse of an increasing function is increasing.

Example 2.12. For a three dimensional solid we set $E = A/V$, where A denotes the surface area and V the volume. For example, for a ball $E(r) = (4\pi r^2)/(\frac{4}{3}\pi r^3) = 3/r$, where r denotes the radius. Then $E'(r) < 0$. The same principle holds for other shapes, E decreases as we enlarge the solid without changing its shape. What does this have to do with the size of animals?

Warm blooded animals living in cold climates need to preserve their body temperature. The total amount of heat stored in the body is proportional to the volume, while the heat loss is proportional to the surface area. The ratio of volume to surface area increases as the animal gets larger, so that for warm blooded animals it is of advantage to be large if they live in cold climates. In hot climates they need to give off heat, so that it is of advantage to be small. Natural selection (Darwinism) should favor the larger specimens of a warm blooded species in a cold climate and smaller ones in a hot climate. You can observe this phenomenon in real life.

For cold blooded animals the converse holds. They absorbe heat so that they body reaches a temperature at which they can be active. In cold climates it helps to be small, because then the surface area is relatively large, compared to the volume. In hot climates cold blooded animals can afford to be large, as it is easy to reach and maintain the temperature at which they can be active. The argument is again consistent with real life.

Needless to say, there are other mechanisms to increase the surface area of a body than decreasing its size, and the maintenance of the body temperature is only one factor which influences the size of specimens of a species. Larger animals need more food, are stronger, cannot hide so well, and are often less agile. All of these factors need to be taken into account to determine the optimal size of an animal. $\diamondsuit$

So far we have only discussed examples where we used (1) and (2) of Theorem 2.11. Let us show how to use the conclusion in (3). To apply it we need to determine intervals on which a function does not change signs. We recall a procedure which works well for continuous functions.

Definition 2.13. *Suppose $f(x)$ is a function. We call a point x_0 on the real line* exceptional *if either $f(x_0) = 0$ or $f(x_0)$ is not defined.*

The following result is an immediate consequence of the Intermediate Value Theorem, see Theorem 1.16 on page 8. Expressed casually it says that a continuous function can change signs only at exceptional points.

Proposition 2.14. *Suppose $f(x)$ is continuous and $f(x)$ has no exceptional points in the interval (x_0, x_1). Then $f(x) > 0$ for all points in the interval (x_0, x_1), or $f(x) < 0$ for all points in the interval (x_0, x_1). In particular, if $f(x)$ is positive at one point in the interval, then it is positive at all points in the interval. If $f(x)$ is negative at one point in the interval, then it is negative at all points in the interval.*

Example 2.15. For example, consider the function

$$f(x) = \frac{x^2(x^2-4)}{x^2+2x-15} = \frac{x^2(x-2)(x+2)}{(x-3)(x+5)}.$$

The zeros of the numerator, and with this the zeros of $f(x)$, are $x = 0$, $x = 2$, and $x = -2$. The zeros of the denominator, i.e., the points where $f(x)$ is not defined, are $x = 3$ and $x = -5$.

According to the proposition, the sign of $f(x)$ remains unchanged on each of the intervals $(-\infty,-5)$, $(-5,-2)$, $(-2,0)$, $(0,2)$, $(2,3)$ and $(3,\infty)$. Counting signs of the factors in the expression for $f(x)$, we see $f(x)$ is positive on the interval $(-\infty,-5)$, negative on $(-5,-2)$, positive on $(-2,0)$ and on $(0,2)$, negative on $(2,3)$, and positive on $(3,\infty)$. You see that the sign changes at some, but not all, exceptional numbers. ♢

Exercise 15. Find intervals on which the following functions do not change signs. Decide whether the functions are positive or negative on these intervals.

$$(1)\ f(x) = x^3 - x^2 - 5x - 3 \quad (2)\ g(x) = \frac{x}{x^3 + 5x^2 - 4x - 20}.$$

We are ready to discuss the monotonicity of functions whose derivative vanishes at some points.

Example 2.16. Find intervals of monotonicity for the function

$$f(x) = 3x^2 + 5x - 4.$$

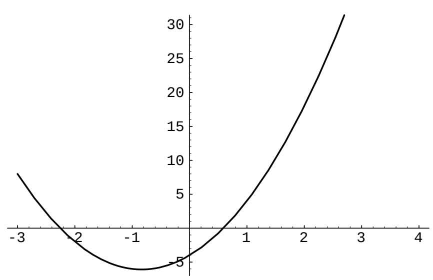

Figure 2.1: A quadratic polynomial, $f(x) = 3x^2 + 5x - 4$

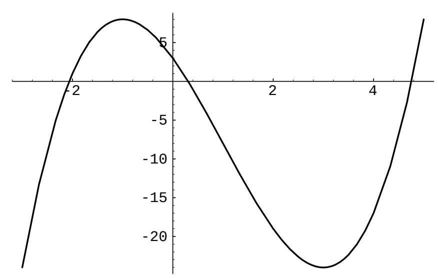

Figure 2.2: A cubic polynomial, $p(x) = x^3 - 3x^2 - 9x + 3$

Solution: We graphed the function in Figure 2.1. Its derivative is $f'(x) = 6x + 5$. In particular, $f'(x) > 0$ if $x \in (-5/6, \infty)$. So $f'(x) > 0$ for all points $x \in [-5/6, \infty)$, except at $x = -5/6$. Theorem 2.11 (3) says that f is increasing on the interval $[-5/6, \infty)$. By a similar argument, f is decreasing on the interval $(-\infty, -5/6]$. ♢

Example 2.17. Find intervals of monotonicity for the degree three polynomial (for a graph see Figure 2.2)

$$p(x) = x^3 - 3x^2 - 9x + 3$$

Solution: The function is defined and differentiable on the real line. Its derivative is

$$p'(x) = 3x^2 - 6x - 9 = 3(x^2 - 2x - 3) = 3(x - 3)(x + 1).$$

Counting the signs of the factors we see that $p'(x)$ is positive on $(-\infty, -1)$ and on $(3, \infty)$. We conclude that $p(x)$ is increasing on the interval $[3, \infty)$ and that it is increasing on the interval $(-\infty, -1]$. The derivative is negative on the interval $(-1, 3)$. The theorem implies that $p(x)$ is decreasing on the interval $[-1, 3]$. $\diamondsuit$

Example 2.18. Find intervals of monotonicity for the rational function

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

Solution: The simplified expression for the derivative of f is

$$f'(x) = \frac{(x + 1)(x - 3)}{(x - 1)^2}.$$

We see that the exceptional points for $f'(x)$ are $x = 1$, $x = -1$ and $x = 3$. We conclude that $f'(x)$ does not change signs on the intervals $(-\infty, -1)$, $(-1, 1)$, $(1, 3)$, and $(3, \infty)$. Counting the signs of the factors of $f'(x)$, we conclude that $f'(x) > 0$ on the intervals $(-\infty, -1)$ and $(3, \infty)$, and $f'(x) < 0$ on the intervals $(-1, 1)$ and $(1, 3)$. Observe that $f(x)$ is defined and differentiable on the entire real line with the only exception of $x = 1$. We conclude that $f(x)$ is increasing on the $(-\infty, -1]$ and $[3, \infty)$. The function is decreasing on the intervals $[-1, 1)$ and $(1, 3]$. $\diamondsuit$

Example 2.19. Find intervals on which the function

$$f(x) = \sin 2x + 2 \sin x$$

is monotonic. Restrict your discussion to the interval $[0, 2\pi]$.

Solution: We differentiate the function and rewrite the expression for the derivative so that it is easier to find its exceptional points.

$$\begin{aligned} f'(x) &= 2\cos 2x + 2\cos x \\ &= 2[2\cos^2 x + \cos x - 1] \\ &= 4(\cos x + 1)\left(\cos x - \frac{1}{2}\right). \end{aligned}$$

To see the second equality we used that $\cos 2x = 2\cos^2 x - 1$. Then we solved the quadratic equation in terms of $\cos x$. We find exceptional points where $\cos x = -1$ (i.e., $x = \pi$) and where $\cos x = \frac{1}{2}$ (i.e., $x = \frac{\pi}{3}$ and $x = \frac{5\pi}{3}$).

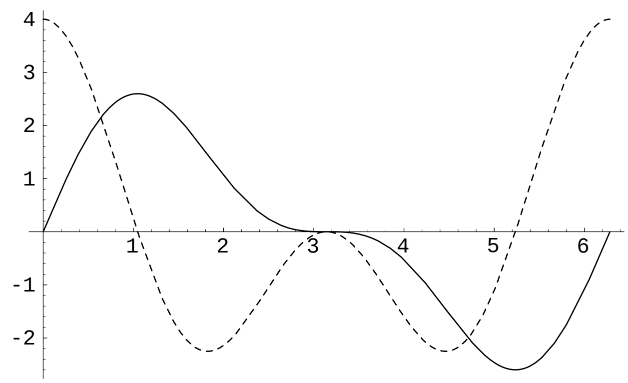

Figure 2.3: A function and its derivative.

Observe that f is differentiable on $[0, 2\pi]$, and that $f'(x) \neq 0$ at the end points of this interval. This provides us with the intervals $[0, \pi/3)$, $(\pi/3, \pi)$, $(\pi, 5\pi/3)$ and $(5\pi/3, 2\pi]$ on which f' does not change sign. Checking the sign of f' (at one point) in each of the intervals, we find that $f'(x) > 0$ for $x \in [0, \pi/3)$ and $x \in (5\pi/3, 2\pi]$, and $f'(x) < 0$ for $x \in (\pi/3, \pi)$ and $(\pi, 5\pi/3)$. We conclude that f is increasing on the interval $[0, \pi/3]$ and $[5\pi/3, 2\pi]$. The function is decreasing on the interval $[\pi/3, 5\pi/3]$, and in this interval there are three points at which $f'(x)$ is not positive.

You may confirm the calculation by having a look at Figure 2.3. There you see the graph of the function (solid line) and the graph of its derivative (dashed line). As you see, wherever $f'(x)$ is positive, there $f(x)$ is increasing. Wherever $f'(x)$ is negative, there $f(x)$ is decreasing. $\diamondsuit$

Exercise 16. Find intervals on which the function f increases and intervals on which f decreases. In the last two problems, (g) and (h), restrict yourself to the interval $[0, 2\pi]$.

(a) $f(x) = 3x^2 + 5x + 7$

(b) $f(x) = x^3 - 3x^2 + 6$

(c) $f(x) = (x + 3)/(x - 7)$

(d) $f(x) = x + 1/x$

(e) $f(x) = x^3(1 + x)$

(f) $f(x) = x/(1 + x^2)$

(g) $f(x) = \cos 2x + 2\cos x$

(h) $f(x) = \sin^2 x - \sqrt{3}\sin x$

2.3.2 Monotonicity at a Point

It is quite natural to ask what it means that a function is increasing at a point, and how this concept is related to the one of being increasing on an interval. We address both questions in this subsection.

Definition 2.20. *Suppose f is a function and c is an interior point of its domain. We say that f is* increasing *at c if, for some $d > 0$,*

$$f(x) < f(c) \text{ for all } x \in (c - d, c) \text{ and } f(x) > f(c) \text{ for all } x \in (c, c + d).$$

We say that f is decreasing *at c if this statement holds with the inequalities reversed.*

Expressed informally, to the left of c the function is smaller and to the right of c it is larger than at c, at least for a while.

Being increasing or decreasing at a point c is a *local* property. We are making a statement about the behavior of the function on some open interval which contains c. Being increasing on an interval is a *global* property. For the global property the interval is given to us. For the local property we may chose the, possibly rather small, interval. The global property has to hold for any two points in the given interval. For the local property we compare $f(x)$ to $f(c)$ where c is fixed and x is any point in an open interval around c which we may chose.

Theorem 2.21. *Suppose f is a function which is defined on an open interval I. Then f is increasing (decreasing) on I if and only it it is increasing (decreasing) at each point in I.*

This theorem establishes the relation between the local and the global property. The 'only if' part is not difficult to show, but the 'if' part uses some deeper facts about finite closed intervals. Our second result gives us a valuable tool to detect monotonicity of functions at a point.

Proposition 2.22. *Let f be a function and c an interior point of its domain. If f is differentiable at c and $f'(c) > 0$, then f is increasing at c. If $f'(c) < 0$, then f is decreasing at c.*

Remark 5. A function does not have to be differentiable to be increasing. Graph the function $f(x) = 2x + |x|$ to convince yourself of this fact. A function can be differentiable and increasing at a point x, even if the assumptions of Proposition 2.22 do not hold, i.e., $f(x) = x^3$ is increasing at $x = 0$, but if $f'(0) = 0$. A function can also be increasing at a point x, but there is not open interval which contains x such that the function is increasing on this interval.

Remark 6. The ideas of of a function being increasing or decreasing at a point may be generalized to cover domains of functions which are half-closed or closed intervals, and where we like to make a statement about the behavior of a function at an endpoint. We have no specific needs for such statements, but the motivated reader is encouraged to explore them.

2.4 The Second Derivative and Concavity

We like to capture the property of a graph being bent upwards or downwards. Secant lines will either be required to lie above or below the graph, and the rates of change will be either increasing of decreasing. These properties can be described globally over intervals and locally at points. You may use the graphs in Figures 2.4 and 2.5 as illustrations of the discussion.

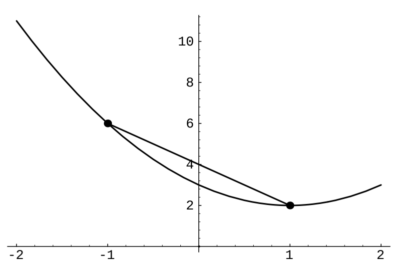

Figure 2.4: Concave Up

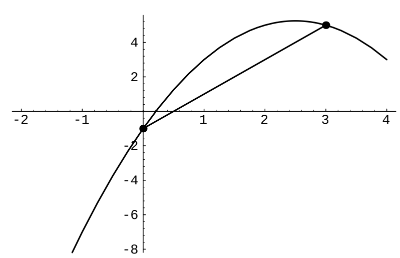

Figure 2.5: Concave Down

2.4.1 Concavity on Intervals

Let $f(x)$ be a function and let $(a, f(a))$ and $(b, f(b))$ be two distinct points on its graph. The line through these two points is

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

If we restrict $l(x)$ to $x \in [a, b]$, then we get the *secant line* through the two point, i.e., the line segment joining the two points.

Definition 2.23. *Let f be a function which is defined on an interval I. We say that f is* concave up *on I if $f(c) < l(c)$ for all a, b in I and $c \in (a, b)$. Here $l(x)$ is the secant line through $(a, f(a))$ and $(b, f(b))$. The inequality expresses that between the points a and b the secant line lies above the graph. We say that f is* concave down *on I if $f(c) > l(c)$ for all a, b in I and $c \in (a, b)$. The inequality expresses that between the points a and b the secant line lies below the graph.*

We state a theorem which provides you with assumptions under which a function is concave up or down. We will not provide a proof of the theorem.

Theorem 2.24. *Let f be a function which is defined on an interval I.*

1. *Suppose that $f(x)$ is differentiable on I. If $f'(x)$ is increasing on I, then $f(x)$ is concave up on I. If $f'(x)$ is decreasing on I, then $f(x)$ is concave down on I.*

2. *Suppose that $f(x)$ is twice differentiable*[3] *on I. If $f''(x) > 0$ for all x in I, then $f(x)$ is concave up on I. If $f''(x) < 0$ for all x in I, then $f(x)$ is concave down on I.*

3. *More generally, the conclusions in (2) still hold if in each finite interval $J \subset I$ there are only finitely many points at which the assumption $f''(x) > 0$, resp. $f''(x) < 0$, is not satisfied.*

For example, the function shown in Figure 2.4 is $q(x) = x^2 - 2x + 3$. Its second derivative is $q''(x) = 2 > 0$. Theorem 2.24 (2) says that q is concave

[3] Strictly speaking, so far we can consider being 'twice differentiable' only for functions which are defined on open intervals. More generally, we proceed as in Section 1.11. We say that $f(x)$ is twice differentiable on I, if $f(x)$ extends to a function $F(x)$ which is defined on an open interval J which contains I, and $F(x)$ is twice differentiable on J. The second derivative will be unique at all points in I if I is not empty and does not consist of exactly one point.

up on $(-\infty, \infty)$. The function shown in Figure 2.5 is $g(x) = -x^2 + 5x - 1$, and its second derivative is $g''(x) = -2 < 0$. Theorem 2.24 (2) says that q is concave down on $(-\infty, \infty)$.

The function $\ln x$ is concave down on the interval $(0, \infty)$. To see this, you may use that $\ln''(x) = -1/x^2 < 0$ on $(0, \infty)$ and apply Theorem 2.24 (2). Alternatively, you may note that the derivative $\ln' x = 1/x$ is decreasing on $(0, \infty)$ and apply Theorem 2.24 (1). The exponential function $\exp(x) = e^x$ is concave up on $(-\infty, \infty)$. To see this, you may note that $\exp''(x) = \exp(x) > 0$ and apply Theorem 2.24 (2). You may also use that $\exp'(x)$ is increasing on the real line, and then quote Theorem 2.24 (1) to derive the desrired conclusion. Finally, you may observe that a function is concave up if its inverse is convave down[4]. So, $\ln x$ being concave down implies that $\exp(x)$ is concave up.

Let us look at examples where we apply condition Theorem 2.24 (3).

Example 2.25. Study the concavity properties of the function

$$p(x) = x^3 - 3x^2 - 9x + 3.$$

Solution: You find the graph of this function in Figure 2.2. Its second derivative is $p''(x) = 6x - 6 = 6(x-1)$. We see that $p''(x) > 0$ for $x \in (1, \infty)$, and $p''(x) < 0$ for $x \in (-\infty, 1)$. This means that $p''(x) > 0$ for all $x \in [1, \infty)$ with only one exception, $x = 1$. Theorem 2.24 (3) tells us that $p(x)$ is concave up on the interval $[1, \infty)$. Similarly, $p''(x) < 0$ for $x \in [-\infty, 1)$ with only one exception, $x = 1$. One deduces that $f(x)$ is concave down on the interval $(-\infty, 1]$. $\Diamond$

Consider the function $\tan x$. You may verify that $\tan'' x = 2\sec^2 x \tan x$. In particular, $\tan'' x < 0$ for $x \in (-\pi/2, 0)$ and $\tan'' x > 0$ for $x \in (0, \pi/2)$. Theorem 2.24 (3) implies that $\tan x$ is concave down on $(-\pi/2, 0]$ and concave up on $[0, \pi/2)$. You may confirm these statements visually by inspecting a graph of the tangent function. You are invited to study the concavity of the other trigonometric and hyperbolic functions.

Remark 7. You may consider the spread of a desease. Denote the number of infected people by $I(t)$. It may be scary if $I'(t) > 0$, i.e., $I(t)$ increases. It is worse, and often true in the early stages of an epedemic, if $I''(t) > 0$.

---

[4]If f and g are inverses of each other, then the graph of one of the functions is obtained from the one of the other one by reflection at the diagonal $x = y$. In this process, secant lines which are above the graph turn into secant lines below the graph. Thus, if f is concave up, then g is concave down, and vice versa.

This means that $I'(t)$ increases, and the desease spreads at an increasing rate. Medical professional will not necessarily wait for the time when $I(t)$, the number of infected people, starts decreasing. When $I''(t)$ turns negative, then $I'(t)$ decreases. The spread or the desease slows. One may hope that eventually $I'(t)$ becomes negative, so that the actual number of sick people decreases. The point at which $I''(t)$ changes signs from being positive to being negative may be considered the turning point in the spread of the desease. One of the recent presidents was confused by a subtle argument of this kind[5].

Let us look at this phenomena in a concrete example. Earlier we considered the logistic equation

$$y' = ay - by^2.$$

See Example 1.75 and the graph of a solution of this diferential equation in Figure 1.15. Use implicit differentiation to find the second derivative:

$$y'' = ay' - 2byy' = (a - 2by)y'.$$

We see that $y'' = 0$ if $y' = 0$ or $y = a/(2b)$. The first case occurs if $y = 0$ or $y = a/b$. We called $y = a/b$ the carrying capacity of the system, and it was the stable equilibrium point. The inflection occurs when y is half the carrying capacity. As long as y is less than $a/(2b)$, the population grows at an increasing rate. If $a/(2b) < y < a/b$, then growth slows. You see the turning point in the graph in Figure 1.15. For a while the population seems to explode, but after a while it levels off so that it does not exceed a the carrying capacity.

Exercise 17. Find intervals on which the following functions are concave up, resp., concave down.

1. $f(x) = x^3 - 4x^2 + 8x - 7$
2. $g(x) = x^4 + 2x^3 - 3x^2 + 5x - 2$
3. $h(x) = x + 1/x$
4. $i(x) = 2x^4 - x^2$
5. $j(x) = x/(x^2 - 1)$
6. $k(x) = 2\cos^2 x - x^2$ for $x \in [0, 2\pi]$.

[5]During a televised presidential debate, one of the candidates said (see the New York Times from October 8th, 1984, page B6): "*Some of these facts and figures just don't add up. Yes, there has been an increase in poverty but it is a lower rate of increase than it was in the preceding years before we got here. It has begun to decline, but it is still going up.*"

2.4.2 Concavity at a Point

The notion of being concave up or down was defined for functions which are defined on intervals. Still, we got a picture how the function has to look like near a point, and this is the behavior which we like to capture in a definition.

Definition 2.26. *Let f be a function and c an interior point* [6] *of its domain. We say that f is* concave up*, resp.,* concave down*, at c if there exists an open interval I and a line l, called a* support line*, such that $l(c) = f(c)$ and*

$$f(x) > l(x), \text{ resp., } f(x) < l(x),$$

for all $x \in I$ with $x \neq c$.

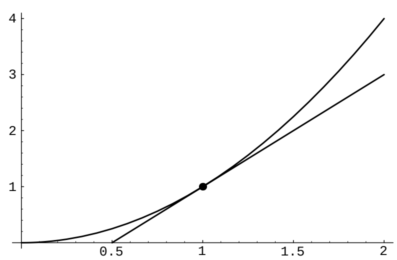

Figure 2.6: Concave up at •

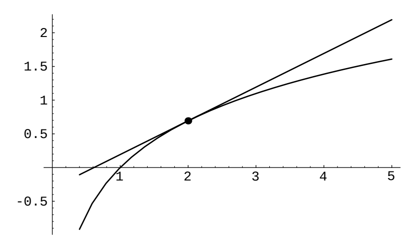

Figure 2.7: Concave down at •

In other words, we are asking for a line $l(x)$, such that the graph lies on one side of the graph, at least near c. If the graph is above the line, then the function is concave up, if it is below, then the function is concave down. We assume that the graph and the line agree at c. You see this situation illustrated in two generic pictures in Figures 2.6 and 2.7. One shows a function which is concave up at the indicated point, one shows a function which is concave down.

Our next theorem tells us how to detect concavity, and it tells us how to find the support line if the function is differentiable.

[6] The idea of an interior point was defined in Definition 1.18 on page 10.

Theorem 2.27. *Let f be a function and c an interior point of its domain.*

1. *If f' is increasing at c or if $f''(c) > 0$, then f is concave up at c.*
2. *If f' is decreasing at c or if $f''(c) < 0$, then f is concave down at c.*
3. *If f is differentiable and concave up or down at c, then there is only one support line, and this line is the tangent line to the graph of f at c.*

The sign of the second derivative of a functions tells us whether a function is concave up or down at a point. If the second derivative is zero, then the test is inconclusive. The function can be concave up, down, or neither. In general, there can be many support lines at any given point, but if the function is differentiable at c, then the support line is unique. It is the tangent line. So, for a differentiable function which is concave up or down at a point, we can draw the tangent line easily. We just hold the ruler against the graph.

For example, the function $f(x) = x^5 - 7x^4 + 2x^3 + 2x^2 - 5x + 4$ is concave down at $x = 2$ because $f''(2) = -148 < 0$.

To relate concavity properties on an interval to those at each point in the interval we state, without proof, the following theorem.

Theorem 2.28. *Let f be a function which is defined on an open interval (a, b). Then f is concave up (resp., down) on (a, b) if and only if f is concave up (resp., down) at each point in (a, b).*

2.5 Local Extrema and Inflection Points

We are going to discuss two types of points which are particularly important in the discussion of (graphs of) functions. As we like to apply local properties of the function, we focus on interior points is the domain of the function.

Definition 2.29 (Local Extrema). *Let f be a function and c an interior point in its domain*[7]*. We say that f has a* local maximum, resp. minimum, at c *if*

$$f(c) \geq f(x), \text{ resp. } f(c) \leq f(x),$$

for all x in some open interval I around c. In this case we call $f(c)$ a local maximum, resp. minimum, of f. A local extremum is a local maximum or minimum.

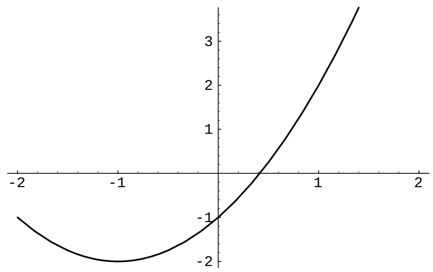

Figure 2.8: A local minimum

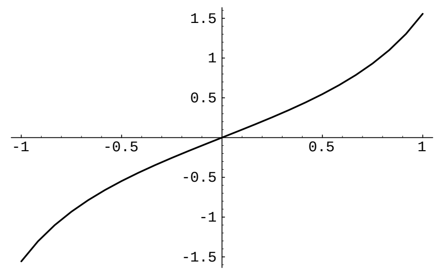

Figure 2.9: An Inflection Point

In other words, f has a local maximum of $f(c)$ at c, if $f(c)$ is the largest value compared the values at points near c. The function shown in Figure 2.8 has a local minimum at $x = -1$. We will study tests which allow us find local extrema soon. We do not need any test to see that $f(x) = |x|$ has a local minimum at $x = 0$, and $f(x) = -(x-1)^2$ has a local maximum at $x = 1$. The vertex of a parabola is always a local extremum, a local minimum if the coefficient of x^2 is positive, and a local maximum if the coefficient of x^2 is negative.

Definition 2.30 (Inflection Points). *Let f be a function and c an interior point of its domain. We call c an* inflection point *of f if the concavity of f changes at c. I.e., for some numbers a and b with $a < c < b$, we have that f is concave up on the interval $(a, c]$ and concave down on $[c, b)$, or vice versa.*

Soon we will develop tests which detect inflections points. No test is required to see that $f(x) = \tan x$ has an inflection point at $x = 0$. The function is concave down on the interval $(-\pi/2, 0]$ and concave up on the interval $[0, \pi/2)$. So the concavity changes at $x = 0$ and that means that there is an inflection point at $x = 0$. You see the graph of this function in Figure 2.9.

[7] According to Definition 1.18 on page 10 this means that $f(x)$ is defined for all x in some open interval around c.

2.6 Detection of Local Extrema

We will discuss how to detect local extrema. The first result excludes many points. Typically, there are very few points where local extrema can occur.

Theorem 2.31. *Let f be a function and c an interior point of its domain. If f is differentiable at c and $f'(c) \neq 0$, then f does not have a local extremum at c. In other words, if f has a local extremum at c, then f is either not differentiable at c or $f'(c) = 0$.*

To have an abbreviation for the points which are recognized as important in this theorem, it is customary to say:

Definition 2.32 (Critical Points). *Let f be a function and c an interior point of its domain. We say that c is a* critical point of f *if f is differentiable at c and $f'(c) = 0$, or if f is not differentiable at c.*

Theorem 2.31 provides us with a necessary condition. If a function has a local extremum at c, then c is a critical point of the function. No local extrema can occur at points which are not critical. The test does not give a sufficient condition for a local extremum. If c is a critical point of the function, then the function need not have a local extremum at c. It makes sense to introduce one more word.

Definition 2.33 (Saddle Points). *Let f be a function and c an interior point of its domain. We say that c is a* saddle point of f *if f is differentiable at c and $f'(c) = 0$, but f does not have a local extremum at c.*

Proof of Theorem 2.31. Suppose that f is differentiable at c and $f'(c) > 0$. Proposition 2.22 on page 76 tells us that there exists some positive number d, such that $f(x) < f(c)$ for all $x \in (c - d, c)$, and $f(x) > f(c)$ for all $x \in (c, c + d)$. So, there are points x to the left of and arbitrarily close to c such that $f(x) < f(c)$, and there are points x to the right of and arbitrarily close to c such that $f(x) > f(c)$. This means, by definition, that f does not have a local extremum at c. If $f'(x) < 0$, then the same argument applies with inequalities reversed. If $f'(c) \neq 0$, then either $f'(c) > 0$ or $f'(c) < 0$, and in neither case we have an extremum at c. □

Neither the exponential function nor the logarithm function have local extrema. To see this, observe that these functions are differentiable on their domain, and their derivatives $\exp' x = \exp x$ and $\ln' x = 1/x$ are everywhere nonzero. These functions have no critical points, and according to Theorem 2.31 they have no local extrema.

Example 2.34. Find the local extrema of the function

$$q(x) = x^2 - 2x + 3.$$

Solution: The function is differentiable for all real numbers x, and

$$q'(x) = 2x - 2 = 2(x - 1).$$

So $q'(x) \neq 0$ if $x \neq 1$. The only point at which we can have a local extremum, i.e., the only critical point, is $x = 1$. If we write the function in the form

$$q(x) = (x - 1)^2 + 2,$$

then we see that q does indeed that a local minimum at $x = 1$. You should confirm this result by having a look at Figure 2.10, where this function is graphed. $\Diamond$

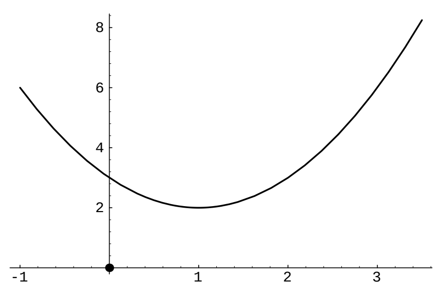

Figure 2.10: A local minimum

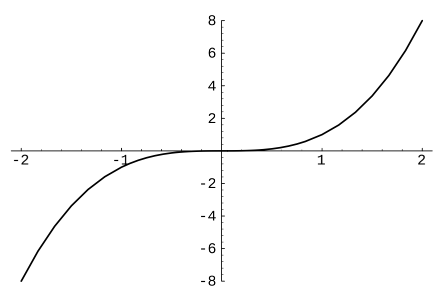

Figure 2.11: A saddle point

Example 2.35. Show that the function $g(x) = x^3$ has a saddle point at $x = 0$.

Solution: The function $g(x)$ is everywhere differentiable, and its only critical point is at $x = 0$, which is the only zero of $g'(x) = 3x^2$. Obviously, $g(x) > 0$ for all $x \in (0, \infty)$ and $g(x) < 0$ for all $x \in (-\infty, 0)$. This means that there is no local extremum at $x = 0$. As $g'(0) = 0$ and there is no local extremum at $x = 0$, the function has a saddle point at this point. This saddle point is shown in Figure 2.11. $\Diamond$

Let us formulate a criterion which confirms that a function has a local extremum at a point c. It gives us a sufficient condition for a local extremum at c.

Theorem 2.36. *Suppose c is an interior point of the domain of a function f, and suppose that for some $d > 0$ the function is increasing on $(c-d, c]$ and decreasing on $[c, c+d)$. Then f has a local maximum at c. If the function is decreasing on $(c-d, c]$ and increasing on $[c, c+d)$, then f has a local minimum at c.*

Taking advantage of the information provided by the first derivative, we obtain the following test.

Theorem 2.37 (First Derivative Test). *Suppose f is a function which is defined and differentiable on $(c-d, c+d)$ for some $d > 0$, and c is a critical point.*

1. *If $f'(x) > 0$ for all $x \in (c-d, c)$ and $f'(x) < 0$ for all $x \in (c, c+d)$, then f has a local maximum at c.*

2. *If $f'(x) < 0$ for all $x \in (c-d, c)$ and $f'(x) > 0$ for all $x \in (c, c+d)$, then f has a local minimum at c.*

3. *If $f'(x) > 0$ for all $x \in (c-d, c) \cup (c, c+d)$, then f has a saddle point at c. This conclusion also holds if $f'(x) < 0$ for all $x \in (c-d, c) \cup (c, c+d)$.*

Let us illustrate the use of the theorem with an example.

Example 2.38. Find the local extrema of the function

$$f(x) = x^3 - 3x^2 + 2x + 2.$$

Solution: We differentiate $f(x)$ and express $f'(x)$ as a product of linear factors:

$$f'(x) = 3x^2 - 6x + 2 = 3\left[x - \left(1 + \frac{\sqrt{3}}{3}\right)\right]\left[x - \left(1 - \frac{\sqrt{3}}{3}\right)\right]$$

It is easy to determine where the factors are zero, positive and negative. We conclude that $f'(x) = 0$ if $x = 1 \pm \sqrt{3}/3$, $f'(x)$ is positive on the intervals $(-\infty, 1 - \sqrt{3}/3)$ and $(1 + \sqrt{3}/3, \infty)$, and $f'(x)$ is negative on the interval $(1 - \sqrt{3}/3, 1 + \sqrt{3}/3)$. You can see graphs of f and f' in Figures 2.12 and 2.13

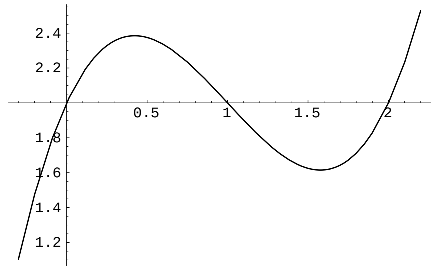

Figure 2.12: $f(x) = x^3 - 3x^2 + 2x + 2$

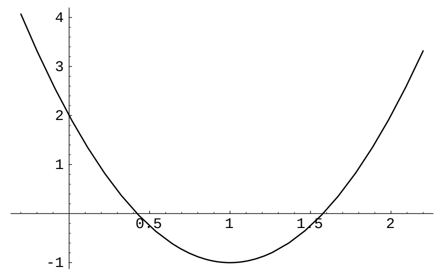

Figure 2.13: $f'(x) = 3x^2 - 6x + 2$

The only only critical points of f are at $x = 1 \pm \sqrt{3}/3$, and these are the only points where a local extremum can occur. Based on the sign of $f'(x)$ on intervals to the left and right of these two critical points we see that f has a local maximum at $x = 1 - \sqrt{3}/3$ and a local minimum at $x = 1 + \sqrt{3}/3$. $\diamondsuit$

Exercise 18. Find the local extrema of the following function:

$$(1)\ f(x) = \frac{x^2 + 3x}{x - 1} \qquad (2)\ g(x) = \sin 2x + 2\sin x \text{ for } x \in [0, 2\pi].$$

Hint: We discussed the monotonicity properties of these functions in Examples 2.18 and 2.19.

Exercise 19. Find the local extrema of the following functions. In the last two problems, (g) and (h), restrict yourself to the interval $[0, 2\pi]$.

(a) $f(x) = 3x^2 + 5x + 7$

(b) $f(x) = x^3 - 3x^2 + 6$

(c) $f(x) = (x + 3)/(x - 7)$

(d) $f(x) = x + 1/x$

(e) $f(x) = x^3(1 + x)$

(f) $f(x) = x/(1 + x^2)$

(g) $f(x) = \cos 2x + 2\cos x$

(h) $f(x) = \sin^2 x - \sqrt{3}\sin x$

Hint: You discussed the intervals of monotonicity for these functions in Exercise 16.

We may use the second derivative to detect the change of sign of the first derivative, as it is called for in the assumptions in Theorem 2.37.

Theorem 2.39 (Second Derivative Test). *Let f be a function and c an interior point in its domain. Assume also that $f'(c)$ and $f''(c)$ exist and that $f'(c) = 0$. If $f''(c) > 0$, then f has a local minimum at c. If $f''(c) < 0$, then f has a local maximum at c.*

To apply the theorem to the detection of the local extrema of a differentiable function $f(x)$, we differentiate f and find the critical points, the zeros of $f'(x)$. Then we differentiate $f'(x)$. The sign of f'' at the critical points tells us whether we found a local minimum or a local maximum. If $f'(c) = f''(c) = 0$, then the test is inconclusive. There may or may not be a local extremum at c. Furthermore, the function f can have a local extremum at c, and the assumptions of the test are not satisfied. In this sense, the test provides us with a sufficient condition for the existence of a local extremum at a point. It does not provide us with a necessary condition.

Example 2.40. Find the local extrema of the function (for a graph, see Figure 2.2 on page 72)

$$p(x) = x^3 - 3x^2 - 9x + 3.$$

Solution: We calculated the first derivative,

$$p'(x) = 3x^2 - 6x - 9 = 3(x + 1)(x - 3).$$

The critical points of the function are $x = -1$ and $x = 3$. Furthermore,

$$p''(x) = 6x - 6 = 6(x - 1).$$

In particular, $p''(-1) = -12$ and $p''(3) = 12$. The second derivative test tells us that we have a local maximum at $x = -1$, because this is a critical point and $p''(-1) < 0$. We also have a local minimum at $x = 3$ because at this critical point the second derivative of the function is positive. ♢

Proof of the Second Derivative Test. First, let us assume that $f'(c) = 0$ and $f''(c) > 0$. We will show that f has a local minimum at c. The assumption that $f'(c) = 0$ means that the tangent line to the graph of f at $(c, f(c))$ is horizontal. Its equation is $l(x) = f(c)$. The assumption that $f''(c) > 0$ means that f is concave up at c (see Theorem 2.27 (1)). Spelled out explicitly this means that

$$f(x) > l(x) = f(c)$$

for some positive number d and for all $x \in (c-d, c) \cup (c, c+d)$. In other words, f has a local minimum at c.

The proof that f has a local maximum at c if $f'(c) = 0$ and $f''(c) < 0$ is similar. We leave it to the reader. □

Exercise 20. Find the critical points and the local extrema.

(a) $f(x) = 4x^2 - 7x + 13$
(b) $f(x) = x^3 - 3x^2 + 6$
(c) $f(x) = x + 3/x$
(d) $f(x) = x^2(1-x)$
(e) $f(x) = |x^2 - 16|$
(f) $f(x) = x^2/(1+x^2)$.

2.7 Detection of Inflection Points

We defined an inflection point to be a point at which the concavity of a function changes. If we know where the function is concave up and down, then we can just answer this question. We want to detect inflection points more efficiently. A theorem provides a necessary and a sufficient condition for the existence of an inflection point. Let us start out with an example.

Example 2.41. Find the the inflection points of the function

$$g(x) = x^3 - 4x^2 + 3x - 5.$$

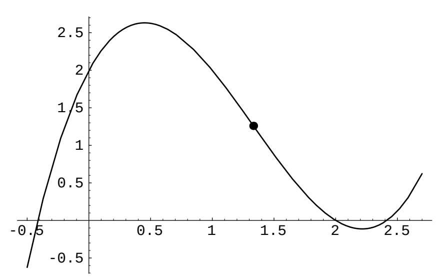

Figure 2.14: The graph of g.

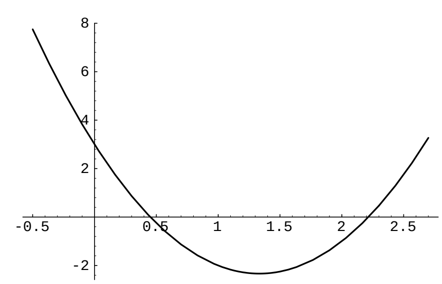

Figure 2.15: The graph of g'.

You see the graph of g in Figure 2.14 and the one of g' in Figure 2.15. We calculate the first and second derivative of g:

$$g'(x) = 3x^2 - 8x + 3 \quad \text{and} \quad g''(x) = 6x - 8.$$

From the formula for the second derivative we conclude that

$$g''(x) < 0 \text{ if } x \in (-\infty, 4/3) \text{ and that } \quad g''(x) > 0 \text{ if } x \in (4/3, \infty).$$

This means that g is concave down on the interval $(-\infty, 4/3]$ and concave up on $[4/3, \infty)$. By definition, we have an inflection point at $x = 4/3$. You see the inflection point indicated as a dot in Figure 2.14. You also see that $g'(x)$ has a local extremum at the same point. $\diamondsuit$

Theorem 2.42. *Let f be a function and c an interior point of its domain. Suppose that the first and second derivatives of f exist at c.*

1. *If f has an inflection point at c, then $f''(c) = 0$.*
2. *If $f''(c) = 0$, $f'''(c)$ exists and $f'''(c) \neq 0$, then f has an inflection point at c.*

Example 2.43. Find the inflection points of

$$f(t) = 2t^4 - 6t^3 + 5t^2 - 7t + 4.$$

Solution: We calculate the second derivative of the function and find

$$f''(t) = 24t^2 - 36t + 10.$$

According to the theorem, we have to find the zeros of $f''(x)$ to determine where an inflection point can be. The roots are

$$t = \frac{3}{4} \pm \frac{1}{12}\sqrt{21} = \frac{9 \pm \sqrt{21}}{12}.$$

Now, let us check whether there are inflection points at either of these values for t. We calculate the third derivative of f:

$$f'''(t) = 48t - 36.$$

We could plug $t = (9 \pm \sqrt{21})/12$ into the expression for f''', but this is a bit cumbersome. We see right away that $f'''(t) = 0$ exactly if $t = 3/4$, and this means that $f'''(9 \pm \sqrt{21})/12) \neq 0$. The theorem says that the inflection points of $f(t)$ are at $t = (9 \pm \sqrt{21})/12$. $\diamondsuit$

Apparently, our ability it find inflection points of a function is limited by our ability to find the zeros of its second derivative. If we are given graphical information, then this quite easy.

Example 2.44. Find the inflection points of the function

$$f(x) = \sqrt{1.2 + x^2 - 3(\sin x)^3}.$$

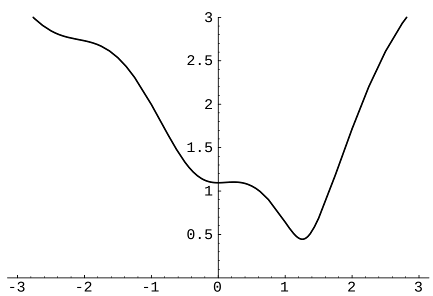

Figure 2.16: The graph of f.

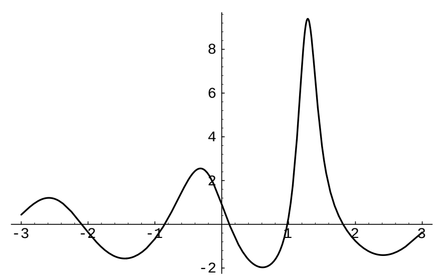

Figure 2.17: The graph of f''.

Apparently, it will take an effort to calculate the second derivative of this function, and it will be nearly impossible to find the zeros of f''. Any reasonable software has no problem with this. We asked the computer to graph f and f'' for $x \in [-3, 3]$. You see the graphs in Figures 2.16 and 2.17.

A look at the graph of f barely reveals some of the inflection points, but the graph of f'' shows them clearly. Zooming in on parts of the graph f will not improve this. At least in this example, the graph of f'' tells us much more about the concavity of the function f than its own graph. $\diamondsuit$

Exercise 21. *Discuss the relation between the inflection points of a function f and the local extrema of its derivative f'.*

2.8 Absolute Extrema of Functions

We said that a function f has a local maximum at c if its value at c is largest in comparison to the values at point near c. In many cases we like to find the maximal value of a function, and where it occurs, anywhere in the domain of the function. This concept is captured in

Definition 2.45. *Let f be a function, and c a point in its domain. We say that f has an* absolute maximum *at c if $f(x) \leq f(c)$ for all x in the domain of f. Then we call $f(c)$ the* absolute maximum *of f. If $f(x) \geq f(c)$ for all x in the domain of f, then we say that f has an* absolute minimum *at c, and we call $f(c)$ the* absolute minimum *of f.*

A different expression is to say that the function assumes its absolute extremum at c.

Theorem 2.46. *A continuous function on a closed interval $[a, b]$ assumes its absolute maximum and minimum either at a critical point or at an endpoint of the interval.*

Proof. In Theorem 1.17 we asserted that a continuous function assumes its absolute maximum at some point in the interval. If the function does not assume its absolute maximum at an endpoint, then it does so at some interior point c, and the function has a local maximum at c. If f is not differentiable at c, the c is critical. If f is differentiable at c, then $f'(c) = 0$ by Theorem 2.31, and c is critical as well. The argument for the absolute minimum is left to the reader. □

Example 2.47. Find the absolute extrema of the function

$$f(x) = x^3 - 5x^2 + 6x + 1$$

for $x \in [0, 4]$.

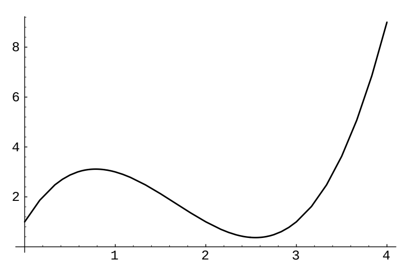

Figure 2.18: $x^3 - 5x^2 + 6x + 1$.

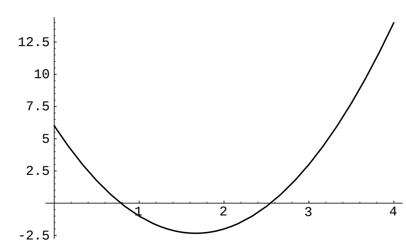

Figure 2.19: $3x^2 - 10x + 6$

Solution: According to Theorem 2.46, the absolute extrema of the function occur either at one of the end points $x = 0$, $x = 4$, or at a critical point. In fact, $f(0) = 1$ and $f(4) = 9$. The critical points, i.e., the zeros of $f'(x) = 3x^2 - 10x + 6$, are $x = (5 \pm \sqrt{7})/3$. Approximate values of these roots are 2.5486 and .7848. You may also check that $f''(x) = 6x - 10$, and

$$f''((5+\sqrt{7})/3) > 0 \quad \text{and} \quad f''((5-\sqrt{7})/3) < 0.$$

The second derivative test tells us that the function has a local minimum at $x = (5+\sqrt{7})/3$ and a local maximum at $x = (5-\sqrt{7})/3$. The approximate values of the function at these points are

$$f((5+\sqrt{7})/3) = 3.1126 \quad \text{and} \quad f((5-\sqrt{7})/3) = .3689.$$

Comparing the values of $f(x)$ at these four points, we conclude that the function assumes its absolute maximum of 9 at $x = 4$, and its absolute minimum of approximately .3689 at $x = (5-\sqrt{7})/3$.

You may compare our calculation with the graphs of f in Figure 2.18 and the one of f' in Figure 2.19. $\diamondsuit$

Exercise 22. Find the absolute extrema of the functions on the indicated intervals.

(a) $f(x) = x^2 - 5x + 2$ for $x \in [0, 5]$

(b) $f(x) = x^3 + 3x^2 - 5x + 2$ for $x \in [-3, 2.5]$

(c) $f(x) = \sqrt{2+x}/\sqrt{1+x}$ for $x \in [0, 5]$

(d) $f(x) = \cos 2x + 2\cos x$ for $0 \le x \le 2\pi$

(e) $f(x) = \sin x + \cos x$ for $0 \le x \le 2\pi$

2.9 Optimization Story Problems

Many real-life problems are formulated as optimization problems. Calculus helps us to solve these optimization problems. To avoid lenghty introductions to real-life problems, we content ourselves with problems of an algebraic or geometric nature. We consider a few examples and give some problems for practice.

Example 2.48. Cut a string of length 50 centimeters into two pieces. Use one piece as the perimeter of an equilateral triangle and the other one as the perimeter of a disk. How long should each piece be, so that the combined area of the triangle and the circle is minimal? How long should each piece be, so that the combined area of the triangle and the circle is maximal?

In our solution we will go through several steps.

Introduction of notation: There are many ways to set up the notation to solve this problem. Among them we say that the side length of the triangle is a and the radius of the circle is r.

Express information as equations: The perimeter of the triangle will be $3a$ and the perimeter of the circle will be $2\pi r$. This means that

$$3a + 2\pi r = 50 \quad \text{and} \quad a = \frac{50 - 2\pi r}{3}.$$

The height of the triangle is $h = \frac{a\sqrt{3}}{2}$, and its area is $\frac{a^2\sqrt{3}}{4}$. The area of the disk is πr^2. The combined area of the triangle and disk is

$$A = \frac{a^2\sqrt{3}}{4} + \pi r^2 = \frac{\sqrt{3}}{4}\left(\frac{50 - 2\pi r}{3}\right)^2 + \pi r^2.$$

For this to make sense, we need that $0 \le r \le 25/\pi$.

Formulate the problem mathematically: Find the absolute minimum (maximum) of the function

$$A(r) = \frac{\sqrt{3}}{4}\left(\frac{50 - 2\pi r}{3}\right)^2 + \pi r^2.$$

for $r \in [0, 25/\pi]$.

Solve the mathematical problem: The derivative of $A(r)$ is

$$A'(r) = -\frac{4\pi}{3} \cdot \frac{\sqrt{3}}{4}\left(\frac{50 - 2\pi r}{3}\right) + 2\pi r = \frac{-\pi}{\sqrt{3}}\left(\frac{50 - 2\pi r}{3}\right) + 2\pi r,$$

and $A'(r) = 0$ if and only if $r = 50/(2\pi + 6\sqrt{3})$. We note that $A(r)$ is a parabola which is open upwards. The critical point, which we just found, is where the local minimum occurs. It is also the absolute minimum of $A(r)$ on any interval which contains the critical point. For the end points we have: $A(0) \approx 120.28$ amd $A(25/\pi) \approx 198.94$.

Answer the original question: The combined area of the disk and the triangle will be minimal if $r = 50/(2\pi + 6\sqrt{3})$, and it will be maximal of $r = 25/\pi$. In the latter case, all string is used for the circle. $\Diamond$

Exercise 23. Repeat the previous example with

1. a disk and a square.

2. an equilateral triangle and a square.

3. a regular hexagon and a square.

4. a disk and half an equilateral triangle (the angles at 30, 60 and 90 degrees).

5. two geometric shapes of your own choice.

Example 2.49. Construct an open box from a rectangular piece of card board of length L and width W. What are the dimensions of the box with the largest possible volume?

In our solution we will go through several steps.

Clarification and introduction of notation: We construct the box by making an incision at a 45 degree angle at each corner. Then we fold up a strip of width x along each side[8]. For yourself, draw a picture of this production process, and convince yourself that any box obtained by a different process will have smaller volume. To simplify matters, we call the longer side of the rectangle L and the shorter one W.

Express information as equations: As we folded up a strip of width x, the box will have width $W-2x$, length $L-2x$, height x, and volume

$$V(x) = (W-2x)(L-2x)x = WLx - 2(L+W)x^2 + 4x^3.$$

By construction, $x \geq 0$, $x \leq W/2$, and $x \leq L/2$, in fact $x \leq W/2$.

Formulate the problem mathematically: Find the absolute maximum of the function

$$V(x) = WLx - 2(L+W)x^2 + 4x^3$$

for $x \in [0, W/2]$.

Solve the mathematical problem: At the end points of the interval V vanishes, i.e., $V(0) = V(W/2) = 0$. On the interior of the interval the function is positive. The derivative of V is

$$V'(x) = WL - 4(W+L)x + 12x^2.$$

The zeros of V' are at

$$x = \frac{1}{6}\left[(L+W) \pm \sqrt{L^2+W^2-LW}\right].$$

[8]You could have cut out a square of size $x \times x$ at each corner.

The function has an inflection point at $(W+L)/6$, to the right of which V'' is positive and V is concave up, and to the left of which V is concave down. We conclude that V has a local maximum at

$$x = \frac{1}{6}\left[(L+W) - \sqrt{L^2 + W^2 - LW}\right].$$

As the function $V(x)$ has only one local maximum in the interval, the local maximum is the same as the absolute maximum.

Answer the original question: The box with the largest volume will have a height of

$$x = \frac{1}{6}\left[(L+W) - \sqrt{L^2 + W^2 - LW}\right].$$

Its width will be $W - 2x$ and its length $L - 2x$. ◊

Exercise 24. Repeat the previous example with specific numbers for the width and length of the piece of card board.

Exercise 25. Start out with an equilateral piece of card board with side length a. Make incisions at the corners, and fold up strips along the edges. You will get an open box whose base is an equilateral triangle. How broad should the folded up strips be, so that the volume of the box is maximal?

Exercise 26. Modify the problem from above, constructing a box with a round base from a circular piece of card board.

Exercise 27. What is the largest possible volume for a right circular cone of slant height a?

Example 2.50. Determine the rectangle of maximal area which can be placed between the x-axis and the graph of the function $f(x) = \sin x$.

Solution: Draw a graph of $\sin x$ so that you can follow the discussion. Convince yourself that the vertices of the rectangle should be $(x, 0)$, $(\pi - x, 0)$, $(x, \sin x)$ and $(\pi - x, \sin x)$ for some $x \in [0, \pi/2]$. The width of the rectangle is $\pi - 2x$ and its height is $\sin x$, so that its area is

$$A(x) = (\pi - 2x)\sin x.$$

We need to find the absolute maximum for this function for $x \in [0, \pi/2]$.

The first derivative of this function is $A'(x) = -2\sin x + (\pi - 2x)\cos x$. After a simple algebraic simplification, you find that

$$A'(x) = 0 \quad \text{if and only if} \quad \tan x = \frac{\pi - 2x}{2}.$$

Find an approximate solution of the equation using Newton's method or your calculator. A fairly good approximation of the zero of $A'(x)$ is $x_0 = .710462$. Convince yourself[9] that this is the only zero of $A'(x)$ for $x \in [0, \pi/2]$. We conclude that x_0 is the only critical point of $A(x)$.

You may calculate $A''(x)$. Substituting x_0 you will see that $A''(x_0) < 0$. It follows from the second derivative test that $A(x)$ has a local maximum at x_0. Apparently $A(x) = 0$ at the end points $x = 0$ and $x = \pi/2$ of the interval. This tells us that $A(x)$ assumes its absolute maximum at x_0.

With this, the final answer to our problem is: The rectangle of maximal area which can be placed between the x-axis and the graph of the sine function will have a width of approximately $\pi - 2x_0 = 1.72066$ and a height of $\sin x_0 = .652183$. Its area will be about 1.12218. ◊

To find the absolute extrema of a continuous function on an interval of the form $[a, b]$ we could inspect the values of the function at the critical points and at a and b. It allows us to decide whether a local extremum is also an abolute one. Our next result allows us to do the same even if the interval is not closed and bounded. The assumptions of this theorem are satisfied in many applied problems.

Theorem 2.51. *Suppose f is defined on an interval I.*

(a) If f is concave up on I and has a local minimum at x_0, then f assumes its absolute minimum at x_0.

(b) If f is concave down on I and has a local maximum at x_0, then f assumes its absolute maximum at x_0.

Example 2.52. Find the absolute minimum of the function

$$f(x) = x + \frac{1}{x}$$

for $x \in (0, \infty)$.

Solution: We calculate the first and second derivative of $f(x)$:

$$f'(x) = 1 - \frac{1}{x^2} \quad \text{and} \quad f''(x) = \frac{2}{x^3}.$$

We find that $f'(x) = 0$ if $x = 1$, and that $f''(x) > 0$ for all x in $(0, \infty)$. So f has a local minimum at $x = 1$, and f is concave up on $(0, \infty)$. Theorem 2.51 tells us that the absolute minimum of the function is $f(1) = 2$. ◊

[9]One possible argument is that $\tan x$ is increasing on the interval $[0, \pi/2)$, and that $\frac{\pi - 2x}{2}$ is decreasing. So these functions can intersect in only one point.

Exercise 28. Find the largest possible area for a rectangle with base on the x-axis and upper vertices on the curve $y = 4 - x^2$.

Exercise 29. A rectangular warehouse will have 5000 m^2 of floor space and will be separated into two rectangular rooms by an interior wall. The cost of the exterior walls is \$ 1,000.00 per linear meter and the cost of the interior wall is \$ 600.00 per linear meter. Find the dimensions of the warehouse that minimizes the construction cost.

Exercise 30. One side of a rectangular meadow is bounded by a cliff, the other three sides by straight fences. The total length of the fence is 600 meters. Determine the dimensions of the meadow so that its area is maximal.

Exercise 31. Draw a rectangle with one vertex at the origin $(0, 0)$ in the plane, one vertex on the positive x-axis, one vertex on the positive y-axis, and one vertex on the line $3x + 5y = 15$. What are the dimensions of a rectangle of this kind with maximal area?

Exercise 32. Two hallways, one 8 feet wide and one 6 feet wide, meet at a right angle. Determine the length of the longest ladder that can be carried horizontally from one hallway into the other one.

Exercise 33. Inscribe a right circular cylinder into a right circular cone of height 25 cm and radius 6 cm. Find the dimensions of the cylinder if its volume is the be a maximum.

Exercise 34. A right circular cone is inscribed in a sphere of radius R. Find the dimensions of the cone if its volume is to be maximal.

Exercise 35. Find the dimensions of a right circular cone of minimal volume, so that a ball of radius 10 centimeters can be inscribed.

Exercise 36. Consider a triangle in the plane with vertices $(0, 0)$, $(a, 0)$, and $(0, b)$. Suppose that a and b are positive, and that $(2, 5)$ lies on the line through the points $(a, 0)$, and $(0, b)$. What should the slope of the line be, so that the area of the triangle is minimal?

Exercise 37. Minimize the cost of the material needed to make a round drum with a volume of 200 liter (i.e., .2 m^3) if

(a) the drum has a bottom and a top, and the same material is used for the top, bottom and sides.

(b) the drum has no top (but a bottom) and the same material is used for the bottom and sides.

(c) the drum has a bottom and a top, the same material is used for the top and bottom, and the material for the top and bottom is twice as expensive as the material for the sides.

(d) the situation is as in the previous case, but the top and the bottom are cut out of squares, and the left over material is recycled for half its value.

Exercise 38. Consider a box with a round base and no lid whose interior is subdivided into six wedge shaped sectors. Which shape should it have, so that its volume is maximal, assuming you are allowed a fixed amount of material? More specifically determine the ratio of radius and height which will maximize the volume.

Exercise 39. Design a roman window with a perimeter of 4 m which admits the largest amount of light. (A roman window has the shape of a rectangle capped by a semicircle.)

Exercise 40. A rectangular banner has a red border and a white center. The width of the border at top and bottom is 15 cm, and along the sides 10 cm. The total area is 1 m^2. What should be the dimensions of the banner if the area of the white area is to be maximized?

Exercise 41. A power line is needed to connect a power station on the shore line to an island 2 km off shore. The point on the coast line closest to the island is 6 km from the power station, and, for all practical purposes, you may suppose that the shore line is straight. To lay the cable costs \$40,000 per kilometer under ground and \$70,000 under water. Find the minimal cost for laying the cable.

Exercise 42. Consider the distance $D(x)$ between a point $P(x) = (x, f(x))$ on the graph of a differentiable function $f(x)$ and a point $Q = (x_0, y_0)$ not on this graph. Suppose $D(x)$ has a local minimum at x_1. Then the tangent line to the graph of f at x_1 intersects the line joining $P(x_1)$ and Q perpendicularly.

2.10 Sketching Graphs

The techiques which we developed so far provide us with some valuable tools for graphing functions. Let us make a list of data which we may determine, so that we can sketch a graph rather precisely. Going through the following program is also a good review of the material which we developed in this chapter.

Useful information for graphing a function: We call the function $f(x)$.

(a) Plot some points on the graph, such as the y-intercept. If the function is given on a closed interval, plot the values at its endpoints.

(b) Plot the zeros of the function. If you cannot find the zeros by analytical means, try it numerically (Newton's method).

(c) If possible, decide on which intervals the function is positive, resp., negative.

(d) Find the first derivative $f'(x)$ of $f(x)$.

(e) Repeat (b) and (c) with $f'(x)$ in place of $f(x)$. Intervals on which $f'(x)$ is positive give you intervals on which $f(x)$ is increasing, and intervals on which $f'(x)$ is negative give you intervals on which $f(x)$ is decreasing. The zeros of $f'(x)$ provide you with the critical points of $f(x)$. Plot the critical points (x and y value), and keep track of the intervals on which the function is increasing, resp., decreasing.

(f) Find the second derivative $f''(x)$ of $f(x)$.

(g) Repeat (b) and (c) with $f''(x)$ in place of $f(x)$. Intervals on which $f''(x)$ is positive give you intervals on which $f(x)$ is concave up, and intervals on which $f''(x)$ is negative give you intervals on which $f(x)$ is concave down. Find the inflection points of the function, i.e., the points where the concavity changes. Plot the inflection points (x and y value), and keep track of the intervals on which the function is concave up, resp., concave down.

(h) Decide at which critical points of $f(x)$ the function has a saddle point or local extremum, and whether it is a minimum or a maximum.

If you now draw a graph which exhibits all of the properties which you gathered in the course of the suggested program, then your graph will look very much like the graph of $f(x)$. More importantly, the graph will have all of the essential features of the graph of $f(x)$. Let us go through the program in an example.

Example 2.53. Discuss the graph of the function

$$f(x) = x^4 - 2x^3 - 3x^2 + 8x - 4 \text{ for } x \in [-3, 3].$$

Solution: To make the discussion a little easier, we note that

$$f(x) = (x-1)^2(x^2-4) = (x-1)^2(x-2)(x+2). \tag{2.1}$$

You should verify this by multiplying out the expression for $f(x)$ in (2.1).

(a): Plot the y intercept of the function and its values at the end points of the given interval: $f(-3) = 80$, $f(0) = -4$ and $f(3) = 20$.

(b): As a polynomial, the function $f(x)$ is differentiable on the given interval. The only exceptional points are its zeros. Having written $f(x)$ as in (2.1), we see right away that $f(x) = 0$ if and only if $x = -2$, $x = 1$, or $x = 2$. Plot these x-intercepts.

(c): Counting the signs of the factors of $f(x)$, we see that $f(x)$ is positive on the intervals $[-3, -2)$ and $(2, 3]$, and negative on $(-2, 1)$ and $(1, 2)$.

(d): We calculate the derivative of $f(x)$:

$$f'(x) = 2(x-1)(x^2-4) + (x-1)^2 2x = 2(x-1)(2x^2 - x - 4).$$

We based the calculation on the description of $f(x)$ in (2.1). In the first step we applied the product rule, and then we used elementary algebra.

(e): We use the quadratic formula to find the zeros of the factor $2x^2 - x - 4$ in the expression for $f'(x)$. They are $(1 \pm \sqrt{33})/4$. This allows us to factor the expression for $f'(x)$, and we find:

$$f'(x) = 4(x-1)\left(x - \frac{1}{4}[1+\sqrt{33}]\right)\left(x - \frac{1}{4}[1-\sqrt{33}]\right).$$

We conclude that:

- $f'(x)$ is negative on the interval $[-3, (1-\sqrt{33})/4)$ and $f(x)$ is decreasing on $[-3, (1-\sqrt{33})/4]$.
- $f'(x)$ is positive on the interval $((1-\sqrt{33})/4, 1)$ and $f(x)$ is increasing on $[(1-\sqrt{33})/4, 1]$.
- $f'(x)$ is negative on the interval $(1, (1+\sqrt{33})/4)$ and $f(x)$ is decreasing on $[1, (1-\sqrt{33})/4]$.
- $f'(x)$ is positive on the interval $((1+\sqrt{33})/4, 3]$ and $f(x)$ is increasing on $[(1+\sqrt{33})/4, 3]$.

- $f(x)$ has a critical point and local minimum at $(1-\sqrt{33})/4 \approx -1.19$, a critical point and local maximum at $x = 1$, and a critical point and local minimum at $(1+\sqrt{33})/4 \approx 1.69$.

The values of the function at its three critical points are approximately:

$$f(\frac{1-\sqrt{33}}{4}) \approx -12.39 \quad \& \quad f(1) = 0 \quad \& \quad f(\frac{1+\sqrt{33}}{4}) \approx -.54.$$

Plot these points.

(f): We rewrite the first derivative as $f'(x) = 4x^3 - 3x^2 - 3x + 4$, and find

$$f''(x) = 12x^2 - 12x - 6.$$

(g): We use the quadratic formula to find the zeros on $f''(x)$ and factor it:

$$f''(x) = 12\left(x - \frac{1}{2}[1+\sqrt{3}]\right)\left(x - \frac{1}{2}[1-\sqrt{3}]\right).$$

We conclude that:

- $f''(x)$ is positive on the interval $[-3, (1-\sqrt{3})/2)$ and $f(x)$ is concave up on $[-3, (1-\sqrt{3})/2]$
- $f''(x)$ is negative on the interval $((1-\sqrt{3})/2, (1+\sqrt{3})/2)$ and $f(x)$ is concave down on $[(1-\sqrt{3})/2, (1+\sqrt{3})/2]$
- $f''(x)$ is positive on the interval $((1+\sqrt{3})/2, 3]$ and $f(x)$ is concave up on $[(1+\sqrt{3})/2, 3]$
- $f(x)$ has inflection points at $x = (1-\sqrt{3})/2 \approx -.37$ and at $x = (1+\sqrt{3})/2 \approx 1.37$.

The values of the function at its inflection points is approximately:

$$f(\frac{1-\sqrt{3}}{2}) \approx -7.21 \quad \& \quad f(\frac{1-\sqrt{3}}{2}) \approx -.29.$$

Plot these points.

(h): At this point we could use the second derivative test to find at which critical points the function has local extrema, but we decided this already based on first derivative behaviour in (e).

Let us gather and organize our information. We consider the interval:

$$I_1 = [-3, -2] \qquad I_5 = \left[1, \frac{1+\sqrt{3}}{2}\right]$$
$$I_2 = \left[-2, \frac{1-\sqrt{33}}{4}\right] \qquad I_6 = \left[\frac{1+\sqrt{3}}{2}, \frac{1+\sqrt{33}}{4}\right]$$
$$I_3 = \left[\frac{1-\sqrt{33}}{4}, \frac{1-\sqrt{3}}{2}\right] \qquad I_7 = \left[\frac{1+\sqrt{33}}{4}, 2\right]$$
$$I_4 = \left[\frac{1-\sqrt{3}}{2}, 1\right] \qquad I_8 = [2, 3]\,.$$

We tabulate the which properties hold on which interval. It should be understood, that at some end points of intervals the function is zero.

Property	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
Sign	pos	neg	neg	neg	neg	neg	neg	pos
Monotonicity	dec	dec	inc	inc	dec	dec	inc	inc
Concavity	up	up	up	down	down	up	up	up

Table 2.1: Properties of the Graph

In Figure 2.20 you see the graph of the function. We have shown it on a slightly smaller interval, as the values at the endpoint a comparetively large. Showing all of the graph would show less clearly what happens near the intercept, extrema, and inflection points. The dots indicate the points which we suggests to plot.

In Figure 2.21 you see the graph of f on an even smaller interval, and parts of the graphs of f' and f''. You can use them to see that f is decreasing where f' is negative, f is concave down where f'' is negative, etc. $\diamondsuit$

Exercise 43. In analogy with the previous example, discuss the function

$$f(x) = (x-1)(x-2)(x+2) = x^3 - x^2 - 4x + 4$$

on the interval $[-3, 2.5]$. In addition, find the absolute extrema of this function.

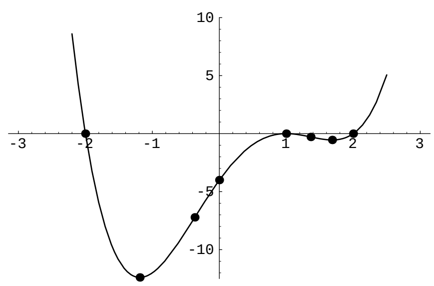

Figure 2.20: The Graph

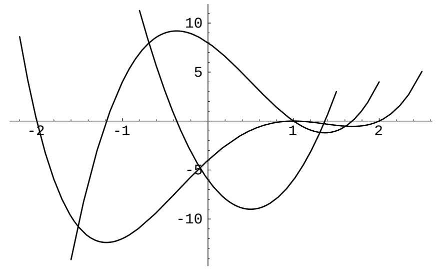

Figure 2.21: f, f', f''

Exercise 44. In analogy with the previous example, discuss the function

$$f(x) = x^3 - 3x + 2$$

on the interval $[-2, 2]$. In addition, find the absolute extrema of this function.

Exercise 45. In analogy with the previous example, discuss the function

$$f(x) = 2\sin x + \cos 3x$$

on the interval $[0, 2\pi]$. In addition, find the absolute extrema of this function. You may have to apply Newton's method to find zeros of f, f', and f''.

Chapter 3

Integration

We will introduce the ideas of the *definite* and the *indefinite integral*. Suppose that f is a function which is defined and bounded on the interval $[a, b]$. If it exists, then the definite integral of f over the interval $[a, b]$ is a real number. It is denoted by

$$\int_a^b f(x)\ dx.$$

The definition is set up, so that for a non-negative function it makes sense to think of the integral as the *area* of the region bounded by the graph of the function, the x-axis, and the lines $x = a$ and $x = b$.

The indefinite integral of a function f is the family (set) of all antiderivatives of f, i.e. all functions whose derivative is f. For important classes of functions one may utilize definite integrals to construct antiderivatives. The Fundamental Theorem of Calculus relates definite integrals and antiderivatives.

To be concrete, consider the function $f(x) = x^2e^{-x}$, shown in Figure 3.1, and find the area of the region Ω bounded by the graph of $f(x)$, the lines $x = 1$ and $x = 5$, and the x-axis.

3.1 Properties of Areas

So far, we only know the area of some simple regions, like rectangles. We will denote the area of a region Ω by Area(Ω). Whatever concept of area we have in mind, it should have the following properties:

- The area of a rectangle is the product of the lengths of its sides.

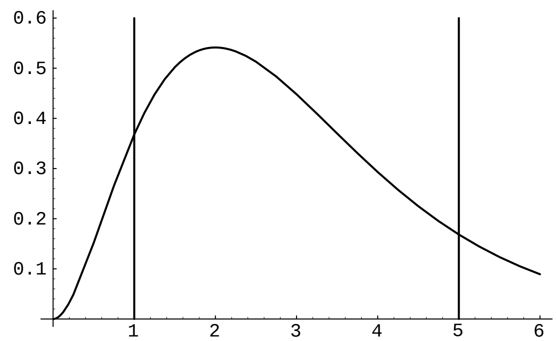

Figure 3.1: $f(x) = x^2 e^{-x}$

- Suppose that Ω_1 and Ω_2 are regions in the plane, and that the area of each of them is defined.

$$\text{If } \Omega_1 \subseteq \Omega_2, \text{ then } \operatorname{Area}(\Omega_1) \leq \operatorname{Area}(\Omega_2).$$

- Suppose that Ω_1 and Ω_2 are regions in the plane, and that the area of each of them is defined. If the regions Ω_1 and Ω_2 do not intersect, then the area of the union $\Omega_1 \cup \Omega_2$ of Ω_1 and Ω_2 is defined, and

$$\operatorname{Area}(\Omega_1 \cup \Omega_2) = \operatorname{Area}(\Omega_1) + \operatorname{Area}(\Omega_2).$$

Suppose for a moment, that the region under the graph shown in Figure 3.1 has an area. In Figure 3.2 you see a rectangle R_l with area .6, which is contained in Ω. In Figure 3.3 you see a rectangle R_u with area 2.24 which contains Ω. The first two principles tell us that

$$\operatorname{Area}(R_l) = .6 \leq \operatorname{Area}(\Omega) \leq \operatorname{Area}(R_u) = 2.24.$$

From above principles one may derive another one, which occurs frequently in our upcoming constructions:

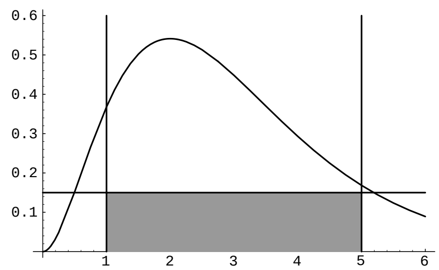

Figure 3.2: A rectangle R_l contained in Ω

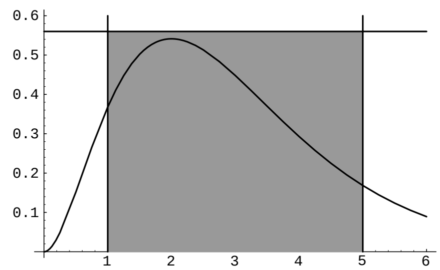

Figure 3.3: A rectangle R_u containing Ω

- Suppose the region R in the plane is the union of a finite number of rectangles $R_1, \ldots, R_n$ and any two of them intersect at most in an edge. Then Area(R) is defined, and it is equal to the sum of the areas of the regions $R_1, \ldots, R_n$:

$$\text{Area}(R) = \text{Area}(R_1) + \cdots + \text{Area}(R_n).$$

3.2 Partitions and Sums

We like to refine the approach to calculating areas of regions which we started in the previous section. We do so by partitioning the interval before applying the ideas from above, and then we add up what we get over the individual intervals.

A *partition* of an interval $[a, b]$ is of a collection is points $\{x_j \mid 0 \leq j \leq n\}$, such that

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

The interval $[a, b]$ is partitioned into n intervals $[x_{j-1}, x_j]$ with $1 \leq j \leq n$.

3.2.1 Upper and Lower Sums

As before, f denotes a function which is defined and bounded on $[a, b]$. On each interval we pick numbers m_j and M_j, such that

$$m_j \leq f(x) \leq M_j \quad \text{for all } x \in [x_{j-1}, x_j].$$

We define the *lower sum* to be

$$S_l = m_1(x_1 - x_0) + m_2(x_2 - x_1) + \cdots + m_n(x_n - x_{n-1}). \tag{3.1}$$

and the *upper sum* to be

$$S_u = M_1(x_1 - x_0) + M_2(x_2 - x_1) + \cdots + M_n(x_n - x_{n-1}). \tag{3.2}$$

These sums depend on the choice of partition and the choices for the m_j and M_j.

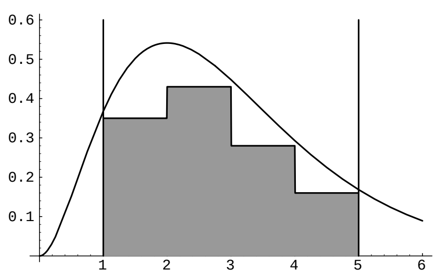

Figure 3.4: A union of rectangles contained in Ω

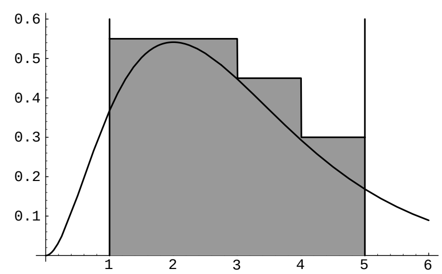

Figure 3.5: A union of rectangles containing Ω

Let us return to the example of the function $f(x) = x^2 e^{-x}$ on the interval $[1, 4]$. In the computation of the lower sum we use the partition

$$x_0 = 1 < x_1 = 2 < x_2 = 3 < x_3 = 4 < x_4 = 5$$

of the interval. We also pick $m_1 = .35$, $m_2 = .43$, $m_3 = .28$ and $m_4 = .16$. This leads to a lower sum $S_l = 1.22$. In the computation of the upper sum we use the partition

$$x_0 = 1 < x_1 = 3 < x_2 = 4 < x_3 = 5$$

of the interval. We also pick $M_1 = .55$, $M_2 = .45$ and $M_3 = .3$. This leads to an upper sum $S_u = 1.85$. The m_j and M_j represent the heights of the rectangles in Figures 3.4 and 3.5, and we trust these figures to show that $m_j \leq f(x)$ and $f(x) \leq M_j$ on the respective interval.

As before, let Ω denote the region under the graph. Then the union of the rectangles shown in Figure 3.4 is contained in Ω, and the union of the rectangles shown in Figure 3.5 contains Ω. Thus, if Ω has an area, the our principles tell us that

$$S_l = 1.22 \leq \text{Area}(\Omega) \leq S_u = 1.85.$$

In fact the only number greater or equal to all lower sums and smaller or equal to all upper sums is $\frac{5}{e} - \frac{37}{e^5}$, and this will be the area of the region Ω. Here e is the Euler number.

Example 3.1. Let us find upper and lower sums for the function

$$f(x) = x^3 - 7x^2 + 14x - 8$$

for $x \in [.5, 4.5]$. In contrast to the function in the previous example, this function is not non-negative.

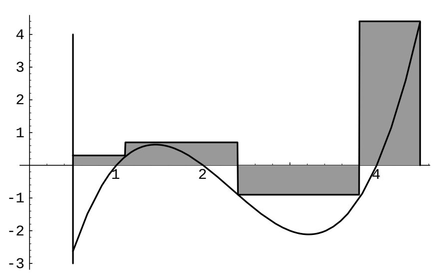

Figure 3.6: Rectangles for calculating an upper sum.

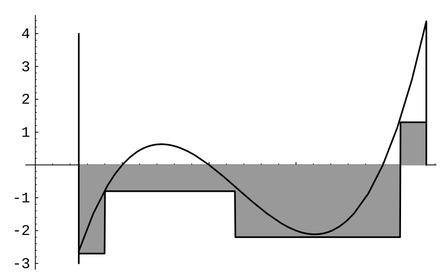

Figure 3.7: Rectangles for calculating a lower sum.

Solution: For the purpose of calculating an upper sum, we partitioned the interval $[.5, 4.5]$ using the intermediate points $x_0 = .5$, $x_1 = 1.1$, $x_2 = 2.4$, $x_3 = 3.8$, and $x_4 = 4.5$. As numbers M_i (so that $M_i \geq f(x)$ for $x \in [x_{i-1}, x_i]$)

we chose $M_1 = .3$, $M_2 = .7$, $M_3 = -.9$, and $M_4 = 4.4$. These data are shown in Figure 3.6. With these choices, the upper sum is

$$\begin{aligned} S_u &= .3(1.1 - .5) + .7(2.4 - 1.1) + (-.9)(3.8 - 2.4) + 4.4(4.5 - 3.8) \\ &= 2.91. \end{aligned}$$

In Figure 3.6 you see four rectangles. Their areas are combined to calculate the upper sum. The areas of the ones above the x-axis are added, the ones below the axis are subtracted, in accordance with the sign of the M_i.

In the calculation of the lower sum we partitioned $[.5, 4.5]$ using $x_0 = .5$, $x_1 = .8$, $x_2 = 2.3$, $x_3 = 4.2$, and $x_4 = 4.5$. As numbers m_i (so that $m_i \le f(x)$ for $x \in [x_{i-1}, x_i]$) we chose $m_1 = -2.7$, $m_2 = -.8$, $m_3 = -2.2$, and $m_4 = 1.3$. These data are shown in Figure 3.7. With these choices we calculate a lower sum of

$$\begin{aligned} S_l &= -2.7(.8 - .5) + (-.8)(2.3 - .8) + (-2.2)(4.2 - 2.3) + 1.3(4.5 - 4.2) \\ &= -5.8. \end{aligned}$$

In Figure 3.7 you see four rectangles. Their areas are combined to calculate the lower sum. The areas of the ones above the x-axis are added, the ones below the axis are subtracted, in accordance with the sign of the m_i.

In summary, you see that we still combine areas of rectangles in the calculation of the upper and lower sum, only that, depending on the sign of the M_i or m_i, these rectangles are either above or below the x-axis, and depending on this, their areas are either added or subtracted. $\Diamond$

Let us make a simple albeit important observation:

Theorem 3.2. *Let f be a function which is defined and bounded on a closed interval $[a, b]$. Let S_l be any lower sum of f and S_u any upper sum. Then*

$$S_l \le S_u.$$

Let us repeat the statement of the theorem to emphasize its meaning. Whichever partition of the interval $[a, b]$ and whichever m_i we use in the calculation of the lower sum S_l and whichever partition of the interval and whichever M_i we use in the calculation of the upper sum S_u, the lower sum is always smaller or equal to the upper sum. To see this, one refines the partitions for the upper and lower sum computation so that they become the same. Then one notes that $m_i \le M_i$ for all i.

3.2.2 Riemann Sums

Suppose once again that $f(x)$ is a function which is defined on the interval $[a, b]$. Pick once more a partition

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

of the interval. In each subinterval, pick a point $\overline{x}_j \in [x_{j-1}, x_j]$. Then we define the *Riemann Sum*

$$S_R = f(\overline{x}_1)(x_1 - x_0) + f(\overline{x}_2)(x_2 - x_1) + \cdots + f(\overline{x}_n)(x_n - x_{n-1}). \tag{3.3}$$

We leave it to the reader to contemplate

Proposition 3.3. *Let f be a function which is defined and bounded on a closed interval $[a, b]$. Let S_l be any lower sum of f, S_u any upper sum, and S_R any Riemann sum. Then*

$$S_l \leq S_R \leq S_u.$$

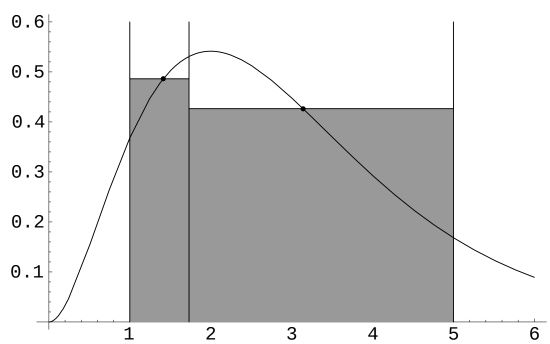

Figure 3.8: Representing a Riemann Sum

To be more concrete, let us return to the example $f(x) = x^2 e^{-x}$ on the interval $[1, 5]$. Let us use the partition

$$x_0 = 1 < x_1 = \sqrt{3} < x_2 = 5.$$

In the two interval of this subdivision we pick the points $\overline{x}_1 = \sqrt{2} \in [1, \sqrt{3}]$ and $\overline{x}_2 = \pi \in [\sqrt{3}, 5]$. As Riemann sum we obtain

$$S_R = f(\overline{x}_1)(x_1 - x_0) + f(\overline{x}_2)(x_2 - x_1) \approx 1.749741.$$

In Figure 3.8 you see the picture illustrating the computation. There are two rectangles, their bases are the intervals in the subdivision, and their heights are $f(\overline{x}_1)$ and $f(\overline{x}_2)$. The sum of the areas of these rectangles is the Riemann sum.

3.3 Limits and Integrability

The idea is to refine the partitions in our previous construction, so that in the limit our sums can be justifiably called the area of the region under the graph, if the function is non-negative. The specifics depend on which sums we are working with.

3.3.1 The Darboux Integral and Areas

As we discussed earlier, whatever choices we make in the calculation of lower and upper sums S_l and S_u, we always have that $S_l \leq S_u$. A crucial additional fact is stated in the next result.

Theorem 3.4. *Let f be a function which is defined and bounded on a closed interval $[a, b]$. There exists a real number Y, such that*

$$S_l \leq Y \leq S_u$$

for all lower sums S_l and upper sums S_u of f.

Idea of Proof. To deduce the theorem from the completeness of the real numbers, one observes that the set of all lower sums of f has a least upper bound. Call it Y_l. The set of all upper sums of f has a greatest lower bound. Call it Y_u. Apparently, $Y_l \leq Y_u$. Then Y is any number such that $Y_l \leq Y \leq Y_u$. □

We are now prepared to define the concept of integrability of a function.

Definition 3.5. *Let f be a function which is defined and bounded on a closed interval $[a, b]$. If there is exactly one number Y, such that*

$$S_l \leq Y \leq S_u$$

for all lower sums S_l and all upper sums S_u of f, then we say that f is integrable *over the interval $[a, b]$. In this case, the number Y is called the* integral[1] *of f for x between a and b. It is also denoted by*

$$\int_a^b f(x)\, dx.$$

[1] To distinguish it from the result of a different, but typically equivalent, construction we should Y the *Darboux integral*.

Remark 8. For completeness sake and later use, let us explain what happens when a function is not integrable. In this case there are at least two different numbers, and with this an entire interval, between all upper and lower sums. So, a function over a closed interval $[a, b]$ is not integrable if and only if the exists a positive number D such that $S_u - S_l \geq D$ for any lower sum S_u and any upper sum S_u.

On the other hand, a function is integrable if for every positive number D there is an upper sum S_u and a lower sum S_l such that $S_u - S_l < D$.

Example 3.6. Explore upper sums, lower sums, and integrability for the function $f(x) = x^2$ on the interval $[0, 1]$.

Solution: Fix a natural number n and set

$$x_0 = 0 < x_1 = \frac{1}{n} < x_2 = \frac{2}{n} < \cdots < x_{n-1} = \frac{n-1}{n} < x_n = \frac{n}{n} = 1.$$

This is an equidistant partition of the interval $[0, 1]$, all subintervals have the same length $1/n$.

For the upper sums we pick

$$M_1 = f(x_1) = \left(\frac{1}{n}\right)^2,\ M_2 = f(x_2) = \left(\frac{2}{n}\right)^2,\ M_3 = f(x_3) = \left(\frac{3}{n}\right)^2,\ \ldots$$

and $M_j = f(x_j) = \left(\frac{j}{n}\right)^2$ in general. Apparently, $M_j \geq f(x)$ for all $x \in [x_{j-1}, x_j]$ because $f(x)$ is increasing on $[0, 1]$. Without proof, we use that

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

We calculate the upper sum

$$\begin{aligned}
S_u &= M_1(x_1 - x_0) + M_2(x_2 - x_1) + \cdots + M_n(x_n - x_{n-1}) \\
&= \left(\frac{1}{n}\right)^2 \times \frac{1}{n} + \left(\frac{2}{n}\right)^2 \times \frac{1}{n} + \cdots + \left(\frac{n}{n}\right)^2 \times \frac{1}{n} \\
&= \frac{1}{n^3}\left[1^2 + 2^2 + \cdots n^2\right] \\
&= \frac{n(n+1)(2n+1)}{6n^3} \\
&= \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}
\end{aligned}$$

For the lower sums we pick

$$m_1 = f(x_0) = 0,\ m_2 = f(x_1) = \left(\frac{1}{n}\right)^2,\ m_3 = f(x_2) = \left(\frac{2}{n}\right)^2,\ \ldots$$

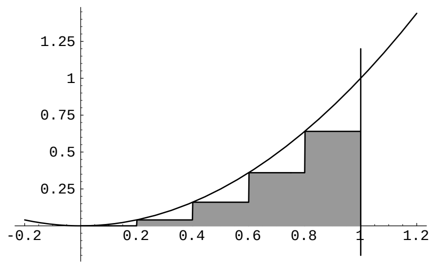

Figure 3.9: Rectangles for calculating a lower sum.

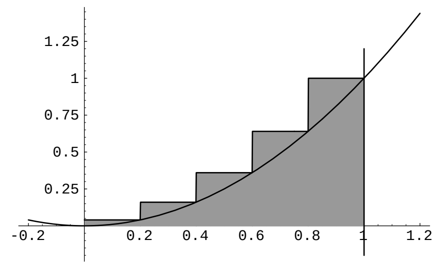

Figure 3.10: Rectangles for calculating an upper sum.

and $m_j = f(x_{j-1}) = \left(\frac{j-1}{n}\right)^2$ in general. The resulting lower sum is

$$S_l = \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}$$

For $n = 5$ you see the rectangles whose areas are the summands in the lower and upper sums in Figures 3.9 and 3.10.

Using the expressions for S_u and S_l you see that $S_u - S_l = 1/n$. We do not only see that $S_l \leq \frac{1}{3} \leq S_u$, but also that $Y = 1/3$ is the only real number, so that $S_l \leq Y \leq S_u$ for all natural numbers n. According to the definition this means, that $f(x) = x^2$ is integrable over the interval $[0, 1]$ and that

$$\int_0^1 x^2 \, dx = \frac{1}{3}. \quad \diamondsuit$$

We motivated our introduction of upper and lower sums by our quest to define the concept of area. Our answer is formulated as a

Definition 3.7. *Let f be a function which is defined, bounded, and non-negative on a closed interval $[a, b]$. Let Ω be the region bounded by the graph of f, the x-axis, and the lines $x = a$ and $x = b$. If f is integrable over this interval, then we say that the region Ω has an* area *and*

$$\text{Area}(\Omega) = \int_a^b f(x) \, dx.$$

The upper and lower sum were constructed such that if there is any justification to assigning an area to Ω then

$$S_l \leq \text{Area}(\Omega) \leq S_u.$$

For an integrable function there is exactly one real number between the lower and upper sums, so this is the only number which we can call the area of Ω.

For example, the area of the region Ω bounded by the graph of the function $f(x) = x^2$, the x-axis, and the lines $x = 0$ and $x = 1$ is

$$\text{Area}(\Omega) = \int_0^1 x^2 \, dx = \frac{1}{3}. \quad \Diamond$$

3.3.2 The Riemann Integral

Earlier we introduced the idea of a Riemann sum. Consider an interval $[a, b]$ and a function $f(x)$ defined on it. We picked a partition

$$\mathcal{P} : a = x_0 \leq x_1 \leq x_2 \leq \cdots x_{n-1} \leq x_n = b,$$

which broke $[a, b]$ up into smaller interval $[x_{j-1}, x_j]$. In each of the subintervals we picked a point $\overline{x}_j \in [x_{j-1}, x_j]$, and set

$$S_R = f(\overline{x}_1)(x_1 - x_0) + f(\overline{x}_2)(x_2 - x_1) + \cdots + f(\overline{x}_n)(x_n - x_{n-1}).$$

We want to consider a *limit Riemann sums*. This is trickier than for functions, because there are a lot of choices which we make to define such a sum. We define the *norm* of the partition $\mathcal{P}$ to be

$$|\mathcal{P}| = \max\{x_j - x_{j-1} \mid 1 \leq j \leq n\},$$

in other words, the norm of $\mathcal{P}$ is the length of the longest of the intervals $[x_{j-1}, x_j]$.

Definition 3.8 (Limit for Riemann Sums). *Suppose the function $f(x)$ is defined on $[a, b]$. We say that*

$$L = \lim_{|\mathcal{P}| \to 0} S_R$$

if for all $\epsilon > 0$ there exists a $\delta > 0$, such that $|L - S_R| < \epsilon$ whenever $|\mathcal{P}| < \delta$. If the limit of the S_R exists, then we say that f is Riemann integrable *over $[a, b]$, call L the* Riemann integral *of f, and write*

$$L = \lim_{|\mathcal{P}| \to 0} S_R = \int_a^b f(x) \, dx.$$

Thus $L = \lim S_R$ if we can force S_R to be close to L, as close as we like, by making the partition fine, by making each subinterval no longer that some number.

It is worth pointing out and not very difficult to show the following proposition.

Proposition 3.9. *Suppose the function f is defined on the interval $[a, b]$. Then f is Riemann integrable if and only if it is Darboux integrable. If defined, the Riemann and the Darboux integral are the same.*

3.4 Integrable Functions

We like to provide a supply of integrable functions. Our first result is typically proved in an analysis course.

Theorem 3.10. *Suppose f is defined and continuous on $[a, b]$. Then f is integrable over $[a, b]$.*

According to this theorem, polynomials are integrable over any interval of the form $[a, b]$. Rational functions (i.e., functions of the form $p(x)/q(x)$ where $p(x)$ and $q(x)$ are polynomials) are integrable over intervals of the form $[a, b]$ as long as q does not vanish anywhere on the interval. The trigonometric functions (sin, cos, tan, cot, sec, and csc) are integrable on intervals where the functions are defined. Arbitrary powers of a variable, $f(x) = x^\alpha$, are integrable. One just needs to make sure that the function is defined on the interval $[a, b]$. For any real number α it suffices to assume that $a > 0$. For any real $\alpha \geq 0$, it suffices to assume $a \geq 0$. For rational numbers $\alpha = p/q$, where p and q are integers and q is odd, it suffices to assume $0 \notin [a, b]$. For non-negative integers α no assumption needs to be made on a and b. Just making sure that the resulting functions are defined everywhere on $[a, b]$, the functions just mentioned may be added, subtracted, multiplied, divided, and composed, and one still ends up with integrable functions.

Let us introduce another class of functions for which we can prove that they are integrable.

Definition 3.11. *Suppose $f(x)$ is a function. We say that $f(x)$ is* non-decreasing *if $f(x_1) \leq f(x_2)$ whenever x_1 and x_2 are in the domain of $f(x)$ and $x_1 \leq x_2$. We say that $f(x)$ is* non-increasing *if $f(x_1) \geq f(x_2)$ whenever $x_1 \leq x_2$.*

Proposition 3.12. *Let $[a, b]$ be a closed interval and let f be defined and non-increasing or non-decreasing on $[a, b]$. Then f is integrable on $[a, b]$. In particular, monotonic (increasing or decreasing) functions are integrable.*

Proof. We will use Darboux integrability. Let us assume that the function f is non-decreasing on the interval. The non-increasing case is left as an exercise. Take any partition of the interval:

$$a = x_0 < x_1 < \cdots < x_n = b.$$

The reader may justify why we can use the same partition in the computation of the upper and lower sum. For $i = 1, \ldots, n$ we set

$$m_i = f(x_{i-1}) \qquad \& \qquad M_i = f(x_i).$$

Then, because f is non-decreasing,

$$m_i \leq f(x) \leq M_i \quad \text{for all } x \in [x_{i-1}, x_i].$$

We use the m_i and M_i to compute upper and lower sums. Let Δ be the largest value of the $x_i - x_{i-1}$. Then

$$\begin{aligned}
S_u - S_l &= [M_1(x_1 - x_0) + \cdots + M_n(x_n - x_{n-1})] \\
&\qquad -[m_1(x_1 - x_0) + \cdots + m_n(x_n - x_{n-1})] \\
&= (M_1 - m_1)(x_1 - x_0) + \cdots + (M_n - m_n)(x_n - x_{n-1}) \\
&\leq [(M_1 - m_1) + (M_2 - m_2) + \cdots + (M_n - m_n)]\,\Delta \\
&= (M_n - m_1)\Delta \\
&= [f(b) - f(a)]\Delta
\end{aligned}$$

The inequality in the computation follows from the choice of Δ. The second to last equality follows because $M_{i-1} = m_i$ for all $i = 2, \ldots, n$. Many terms in the computation cancel. Given any positive number D, we can make the partition fine enough so that $[f(b) - f(a)]\Delta < D$. According to our Remark 8 this means that f is integrable over the interval, as we claimed.

We illustrate the steps in the proof in a concrete example. In Figure 3.11 you see the upper and lower sum. The lower sum is the sum of the areas of the darkly shaded rectangles. The upper sum is the sum of the areas of the lightly and darkly shaded rectangles. The difference between the upper and the lower sum is the sum of the lightly shaded rectangles shown in Figure 3.12. We can combine these areas by sliding the rectangles sideways so that they form one column. Its height will be $f(b) - f(a)$. Its width may vary, but in the widest place it is no wider than Δ, the width of the largest interval in the partition of $[a, b]$. That means, the difference between the upper and the lower sum is at most $[f(b) - f(a)]\Delta$. As above, we conclude that the function is integrable. □

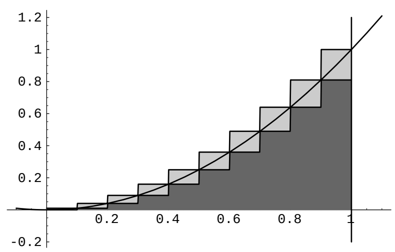

Figure 3.11: Rectangles for calculating a lower and an upper sum.

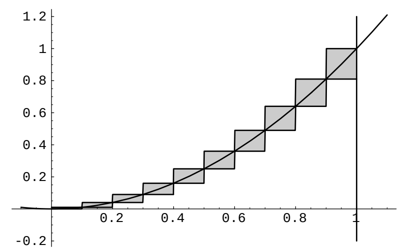

Figure 3.12: Rectangles for calculating the difference between an upper and a lower sum.

Remark 9. There are functions which are not integrable over any interval of the form $[a, b]$ with $a < b$.

Remark 10. Here we only discuss integrability of function over closed finite intervals, i.e., intervals of the form $[a, b]$. The discussion of integrability of functions over intervals which are not of this form, e.g., half-open intervals like $[a, b)$ or unbounded closed intervals like $[a, \infty)$, requires additional ideas and techniques which we are not ready to discuss yet.

3.5 Some elementary observations

In spite of our success calculating some integrals using upper and lower sums and the definition, this is certainly not the way to go in general. To integrate "well behaved" functions we want a theory which allows us to calculate integrals more easily. We have to develop a few basic tools. These are fairly straight forward consequences of the definition of the integral.

Proposition 3.13. *If the function f is defined at a, then*

$$\int_a^a f(x)\, dx = 0 \tag{3.4}$$

Proof. The reader should contemplate the proposition. □

Proposition 3.14. *Let $[a,b]$ be a closed interval, c a point between a and b, and f a function which is defined on the interval. Then*

$$\int_a^c f(x)\ dx + \int_c^b f(x)\ dx = \int_a^b f(x)\ dx. \tag{3.5}$$

Implicitly in the formulation of the proposition is the statement that f is integrable over $[a,b]$ if and only if it is integrable over the intervals $[a,c]$ and $[c,b]$. If one of the sides of Equation (3.5) exists, then so does the other one.

Idea of Proof. Use c as one of the points in the partition. The remaining details are left to the reader. □

As an immediate consequence of Propositions 3.12 and 3.14 we find

Corollary 3.15. *Let f be defined on the interval $[a,b]$. Suppose that we can partition the interval into a finite number of intervals such that f is non-increasing or non-decreasing on each of them. Then f is integrable on $[a,b]$.*

We can also extend Theorem 3.10.

Definition 3.16. *Suppose that f is defined on an interval $[a,b]$. We call f* piecewise continuous *if there is a partition*

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

such that f is continuous on the open intervals (x_{j-1}, x_j) for all $1 \le j \le n$, and the one-sided limits (see Section 1.3)

$$\lim_{x \to x_{j-1}^+} f(x) \quad \textit{and} \quad \lim_{x \to x_j^-} f(x).$$

exist and are finite.

Corollary 3.17. *If f is a piecewise contiuous function on $[a,b]$, then f is integrable on $[a,b]$.*

Idea of Proof. According to Proposition 3.14 we may break the problem up, and consider it over each of the intervals $[x_{j-1}, x_j]$ separately. On each of these smaller intervals, we can change the defintion of the function at a point or two and make it continuous. This changes neither the integrability nor the value of the integral. So the assertion follows from Theorem 3.10. □

Definition 3.18. *Let f be defined and integrable on the interval $[a, b]$. Then*

$$\int_a^b f(x)\ dx = -\int_b^a f(x)\ dx.$$

This definition is convenient and consistent with what we have said so far about the integral. The approach to integrals via lower and upper sums could also be generalized to include integrals $\int_a^b$ where $b < a$, leading to exactly this formula.

Using the definition of the integral it is not difficult to show:

Proposition 3.19. *Let $[a, b]$ be a closed interval and c a scalar. Suppose that f and g are integrable over the interval. Then $f+g$ and cf are integrable over $[a, b]$ and*

$$\int_a^b (f(x) + g(x))\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$$

and

$$\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx.$$

We mention a few useful estimates for integrals.

Proposition 3.20. *If f is integrable over $[a, b]$, and $f(x) \geq 0$ for all $x \in [a, b]$, then*

$$\int_a^b f(x)\ dx \geq 0.$$

Proof. The proof is left to the reader. □

Corollary 3.21. *If h and g are integrable over $[a, b]$, and $g(x) \geq h(x)$ for all $x \in [a, b]$, then*

$$\int_a^b g(x)\ dx \geq \int_a^b h(x)\ dx.$$

Proof. Use that $f(x) = g(x) - h(x) \geq 0$ for all $x \in [a, b]$. □

Proposition 3.22. *Let $[a, b]$ be a closed interval and f integrable over $[a, b]$. Then the absolute value of f is integrable over $[a, b]$, and*

$$\left|\int_a^b f(x)\ dx\right| \leq \int_a^b |f(x)|\ dx. \tag{3.6}$$

The proof of this proposition is elementary, though a bit tricky.

3.6 Areas and Integrals

Let us return to the relation between areas and integrals. Suppose $f(x)$ is a non-negative integrable function over an interval $[a, b]$. If Ω is the area bounded by the graph of $f(x)$, the x-axis, and the lines $x = a$ and $x = b$, then

$$\text{Area}(\Omega) = \int_a^b f(x)\, dx.$$

The question is, what happens if $f(x)$ is not non-negative?

Let f be a function which is defined and bounded on a closed interval $[a, b]$ and Ω the set of points which lie between the graph of $f(x)$ and the x-axis for $a \leq x \leq b$. We decompose Ω into the union of two sets, Ω^+ and Ω^-. Specifically, Ω^+ consist of those points (x, y) in the plane for which $a \leq x \leq b$ and $0 \leq y \leq f(x)$, and Ω^- of those points for which $a \leq x \leq b$ and $f(x) \leq y \leq 0$. Then Ω is the union of the sets Ω^+ and Ω^-. We decompose the region between the x-axis and the graph into the part Ω^+ above the x-axis and the part Ω^- below it. Making use of this notation, we have:

Proposition 3.23. *If f is integrable, then the areas of the regions Ω^+ and Ω^- are defined*[2] *and*

$$\int_a^b f(x)\, dx = \text{Area}(\Omega^+) - \text{Area}(\Omega^-). \tag{3.7}$$

Idea of Proof. We define two functions:

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) \leq 0 \end{cases} \quad \text{and} \quad f^-(x) = \begin{cases} f(x) & \text{if } f(x) \leq 0 \\ 0 & \text{if } f(x) \geq 0 \end{cases}$$

It is elementary, though a bit tricky, to show that the integrability of $f(x)$ implies the integrability of $f^+(x)$ and $f^-(x)$. Apparently, $f = f^+ + f^-$, so that the additivity of the integral implies that

$$\int_a^b f(x)\, dx = \int_a^b f^+(x)\, dx + \int_a^b f^-(x)\, dx. \tag{3.8}$$

According to Definition 3.7 we have

$$\text{Area}(\Omega^+) = \int_a^b f^+(x)\, dx. \tag{3.9}$$

[2]If you want to be formal, then you have to flip the region Ω^- to lie above the x-axis. Only then have we addressed the question of it having an area.

Let $-\Omega^-$ be the area obtained by flipping Ω^- up, i.e., we take its mirror image along the x-axis. This process does not change areas, so $\text{Area}(\Omega^-) = \text{Area}(-\Omega^-)$. The function $-f^-(x)$ is non-negative, and $-\Omega^-$ is bounded by the graph of $-f^-(x)$, the x-axis, and the lines $x = a$ and $x = b$. According to Definition 3.7 and our elementary properties of the integral we have

$$\text{Area}(\Omega^-) = \text{Area}(-\Omega^-) = \int_a^b -f^-(x)\ dx = -\int_a^b f^-(x)\ dx. \tag{3.10}$$

Our claim follows now by substituting the results in (3.9) and (3.10) into (3.8). □

For example, $\int_{-\pi/2}^{\pi/2} \sin x\ dx = 0$ because the graph bounds congruent regions above and below the x-axis.

3.7 Anti-derivatives

Consider a function $f(x)$ with domain I. In Definition 2.6 we called a function $F(x)$ with domain I an antiderivative of $f(x)$ if $F'(x) = f(x)$. Having an anti-derivative of a function will (typically) make it easy to integrate it over a closed interval.

Remember that any antiderivatives F_1 and F_2 of a function f on an interval I differ only by a constant (see Corollary 2.5). In other words, there exists a constant c, such that

$$F_1(x) = F_2(x) + c \quad \text{for all } x \in I.$$

Definition 3.24. *Let f be a function which is defined on an interval I, and suppose that f has an antiderivative. The set of all antiderivatives of f is called the* indefinite integral *of f. It is denoted by*

$$\int f(x)\ dx.$$

Given a function f and an antiderivative F of it, we typically write

$$\int f(x)\ dx = F(x) + c. \tag{3.11}$$

In this expression c stands for an arbitrary constant. Different values for c result in different functions. Allowing all real numbers as possible values for c, we understand the the right hand side of (3.11) as a set of functions. The constant c in the expression is referred to as *integration constant.*

Example 3.25. Given a function $f(x)$ we might know or guess a function $F(x)$, such that $F'(x) = f(x)$. Then we can write down the indefinite integral of f in the form $F(x) + c$. You can check the correctness of your guess by differentiation. You may want to consult Table 1.3 on page 63 to come up with ideas for antiderivatives. Here are some examples.

$$\int 1\ dx = x + c \qquad \int x\ dx = \frac{1}{2}x^2 + c$$

$$\int \sqrt{x}\ dx = \frac{2}{3}x^{3/2} + c \qquad \int x^n\ dx = \frac{1}{n+1}x^{n+1}\ (n \neq -1)$$

$$\int \sin x\ dx = -\cos x + c \qquad \int \sec^2 x\ dx = \tan x + c$$

$$\int \cos x\ dx = \sin x + c \qquad \int \sec x \tan x\ dx = \sec x + c$$

$$\int \frac{dx}{1+x^2} = \arctan x + c \qquad \int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + c$$

Using the linearity of the differentiation (see the differentiation rules in (1.20)), it is easy to produce more examples. E.g.

$$\int 5x^2 - 2\cos x\ dx = \frac{5}{3}x^3 - 2\sin x + c.$$

Occasionally, an additional idea is required before we can see the anti-derivative. E.g., using the trigonometric identity $\cos^2 x = (1 + \cos 2x)/2$, we find that

$$\int \cos^2 x\ dx = \frac{1}{2}\int (1 + \cos(2x))\ dx = \frac{1}{2}\left[x + \frac{1}{2}\sin(2x)\right] + c.$$

Using a different trigonometric identity we find

$$\int (1 + \cot^2 x)\ dx = \int \csc^2 x\ dx = -\cot x + c. \quad \Diamond$$

We shall explore additional ideas for finding antiderivatives at a later. The reader may practice finding some antiderivatives for the functions in the next exercise. As you go through them you are expected to learn, or pick up some new ideas as you go along.

Exercise 46. Find the following indefinite integrals:

(a) $\int 3\ dx$

(b) $\int (x+4)\ dx$

(c) $\int (x^2-5)\ dx$

(d) $\int \cos 2x\ dx$

(e) $\int (3+x)^3\ dx$

(f) $\int (3+2x)^5\ dx$

(g) $\int \frac{1}{x^3}\ dx$

(h) $\int \csc^2 x\ dx$

(i) $\int (1+\tan^2 x)\ dx$

(j) $\int \csc x \cot x\ dx$

(k) $\int \sin^2 x\ dx$

(l) $\int \sec^2(3x)\ dx$

(m) $\int e^{x/3}\ dx$

(n) $\int \frac{2x}{x^2+1}\ dx$

(o) $\int (4-3x)^5\ dx$

(p) $\int \cos(4-3x)\ dx$

(q) $\int \frac{2x}{(x^2+3)^2}\ dx$

(r) $\int x \sec^2(x^2+5)\ dx$

3.8 The Fundamental Theorem of Calculus

Our first result provides us with a large class of functions which have antiderivatives.

Theorem 3.26. *Continuous functions, defined over intervals, have antiderivatives. More specifically, suppose that a function f is defined and continuous over the interval I. Let $a \in I$. Then*

$$f(x) = \frac{d}{dx}\int_a^x f(t)\ dt$$

for all $x \in I$.

The major tool for calculating integrals, and the grand conclusion of our discussion of antiderivatives is the *Fundamental Theorem of Calculus*.

Theorem 3.27 (Fundamental Theorem of Calculus). *Suppose that f is a continuous function over a closed interval $[a, b]$ and that F is an antiderivative of f. Then*

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

For example, $F(x) = -\cos x$ is an antiderivative of $f(x) = \sin x$, so that the Fundamental Theorem of Calculus tells us that

$$\int_0^\pi \sin x\ dx = -\cos(\pi) - (-\cos(0)) = -(-1) - (-1) = 2.$$

As another example, note that $F(x) = \tan x$ is an anti-derivative of $f(x) = \sec^2 x$, so that the Fundamental Theorem of Calculus tells us that

$$\int_0^{\pi/4} \sec^2 x \ dx = \tan(\pi/4) - \tan(0) = 1.$$

Remark 11 (Notational Convention). One commonly uses the notation

$$F(x)\Big|_a^b = F(b) - F(a).$$

This is quite convenient. E.g., we write

$$\sin x\Big|_0^\pi = \sin \pi - \sin 0.$$

If there are ambiguities due to the length of the expression to which this construction is applied, we also use the notation shown in the following example:

$$\Big[x^3 - 5x^2 + 2x - 8\Big]_3^5 = p(5) - p(3)$$

where $p(x) = x^3 - 5x^2 + 2x - 8$.

Using this notation, we calculate that

$$\int_{-2}^{3} (x^2 - 2x + 5) \ dx = \left[\frac{x^3}{3} - x^2 + 5x\right]_{-2}^{3} = \frac{95}{3}.$$

Other examples are

$$\int_0^{\pi/4} \sec x \tan x \ dx = \sec x\Big|_0^{\pi/4} = \sqrt{2} - 1$$

and

$$\int_{\pi/4}^{\pi/3} \csc x \cot x \ dx = -\csc x\Big|_{\pi/4}^{\pi/3} = \left(\frac{-2\sqrt{3}}{3}\right) - (-\sqrt{2}) = \sqrt{2} - \frac{2\sqrt{3}}{3}.$$

The reader is invited to practice a few examples.

Exercise 47. Evaluate the following definite integrals:

(a) $\displaystyle\int_0^1 (3x+2)\ dx$

(b) $\displaystyle\int_1^2 \frac{6-t}{t^3}\ dt$

(c) $\displaystyle\int_2^5 2\sqrt{x-1}\ dx$

(d) $\displaystyle\int_1^0 (t^3-t^2)\ dt$

(e) $\displaystyle\int_{\pi/6}^{\pi/4} \csc x \cot x\ dx$

(f) $\displaystyle\int_{-1}^{-1} 7x^6\ dx$

(g) $\displaystyle\int_0^\pi \frac{1}{2}\cos x\ dx$

(h) $\displaystyle\int_0^\pi \cos(x/2)\ dx$

(i) $\displaystyle\int_{-2}^2 |x^2-1|\ dx$

(j) $\displaystyle\int_0^{\pi/2} \cos^2 x\ dx$

(k) $\displaystyle\int_0^{\pi/2} \sin^2(2x)\ dx$

(l) $\displaystyle\int_0^{\pi/4} \sec^2 x\ dx$

3.8.1 Some Proofs

Because of their importance, we like to prove Theorem 3.26 and the Fundamental Theorem of Calculus.

Proof of the Fundamental Theorem of Calculus. Essentially, the desired result is an easy consequence of Theorem 3.26. Let $F(x)$ be any anti-derivative of $f(x)$ on I, and $H(x) = \int_a^x f(t)\ dt$ the one provided by Theorem 3.26. In particular, $F'(x) = H'(x) = f(x)$. Cauchy's Theorem (see its application in Corollary 2.5) tells us that F and H differ by a constant. For some constant c and all $x \in I$:

$$H(x) = \int_a^x f(t)\ dt = F(x) + c \tag{3.12}$$

We can find out the value for c by substituting $x = a$ in this equation. In particular, we find that

$$\int_a^a f(t)\ dt = 0 = F(a) + c \qquad \text{or} \qquad c = -F(a).$$

Using this calculation of c and substituting $x = b$ in (3.12), we obtain

$$\int_a^b f(t)\ dt = F(b) - F(a),$$

as claimed. $\square$

Proof of Theorem 3.26. Because we assumed continuity of f on the interval I, it follows from Theorem 3.10 that

$$F(x) = \int_a^x f(t)\ dt$$

exists. So it is our task to show that F is differentiable at x, and that $F'(x) = f(x)$. Using Theorem 1.26, after adjusting the notation to fit the current setting, the task becomes to show that

$$\begin{aligned} f(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \left[\int_a^{x+h} f(t)\ dt - \int_a^x f(t)\ dt \right] \\ &= \lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\ dt. \end{aligned}$$

Here we assume that x is not an endpoint of I, so that x and $x+h$ are both in I. We omit (leave to the reader) the modifications of the proof which are required in the case where x is an endpoint of I.

According to the Extreme Value Theorem (see Theorem 1.17) there are points c and d between x and $x+h$, such that

$$f(c) = m(h) \le f(x) \le f(d) = M(h) \tag{3.13}$$

for all t between x and $x+h$. The points c and d may not be uniquely determined by h, but m and M are. It follows from (3.13) and Corollary 3.21 that

$$m(h) \cdot h = \int_x^{x+h} m(h)\ dt \le \int_x^{x+h} f(t)\ dt \le \int_x^{x+h} M(h)\ dt = M(h) \cdot h,$$

and with this that

$$m(h) \le \frac{1}{h} \int_x^{x+h} f(t)\ dt \le M(h).$$

Continuity of $f(x)$ implies that

$$\lim_{h\to 0} m(h) = f(x) = \lim_{h\to 0} M(h).$$

It follows from a pinching argument (see Proposition 1.4) that

$$\lim_{h\to 0} \int_x^{x+h} f(t)\ dt = f(x),$$

and this is exactly what we needed to show. ∎

3.9 Substitution

In some cases it is not that easy to 'see' an antiderivative of the function one likes to integrate. *Substitution* is a method which, when applied correctly, will simplify the expression for the function you like to integrate. You hope that you can find an antiderivative for the simplified expression. The method is based on the chain rule for differentiation. Sometimes this method is helpful, other times it is not. Your success with this method depends greatly on experience, i.e., practice.

We explain the method. Let F and g be functions which are defined and differentiable on an interval I. Set $F' = f$. Then, according to the chain rule,

$$\frac{d}{dx}F(g(x)) = f(g(x))g'(x).$$

Assume that f and g' are continuous on I. Then $f(g(x))g'(x)$ is continuous as well. We may take antiderivatives of both sides of our previous equation, and conclude that

$$\int f(g(x))g'(x)\ dx = F(g(x)) + c. \tag{3.14}$$

The variable for the functions f and F is often called u, and this means in context that $u = g(x)$.

Let us give a few examples to illustrate how this method can be put to use. There are no general rules what substitution must be used, rather success justifies the means. Working through the examples will teach you how to apply this method in some typical situations. It will give you at least some experience which you may then rely on in similar examples.

For example,

$$\int (2x-3)^3\ dx = \frac{1}{2}\int (2x-3)^3 \cdot 2dx = \frac{1}{8}(2x-3)^4 + c.$$

Here we used $g(x) = 2x - 3$, $g'(x) = 2$, $f(u) = u^3$, and $F(u) = \frac{u^4}{4}$.

There is a pattern, a way to use the notation, which can be applied to write down the steps in an integration using substitution efficiently. Setting $u = g(x)$ we write

$$du = g'(x)dx,$$

instead of $g'(x) = du/dx$[3]. Suppose also that F is an anti-derivative of f, so $F' = f$. Then the pattern for calculating an integral via substitution is

$$\int f(g(x))g'(x)\ dx = \int f(u)\ du = F(u) + c = F(g(x)) + c. \tag{3.15}$$

In the first step of this calculation we carry out the substitution, in the second one we find the anti-derivative, and in the third one we reverse the substitution. We make use of this notation in our next example.

For example, we calculate that

$$\begin{aligned}\int x\sqrt{x^2+2}\ dx &= \frac{1}{2}\int \sqrt{x^2+2}\cdot 2xdx\\ &= \frac{1}{2}\int \sqrt{u}\ du\\ &= \frac{1}{3}u^{3/2} + c\\ &= \frac{1}{3}(x^2+2)^{3/2} + c.\end{aligned}$$

We used the substitution $u = x^2 + 2$. Then $\frac{du}{dx} = 2x$, or $du = 2xdx$.

We calculate that

$$\begin{aligned}\int t^2(t+1)^7\ dt &= \int (u-1)^2u^7\ du\\ &= \int (u^2-2u+1)u^7\ du\\ &= \int (u^9-2u^8+u^7)\ du\\ &= \frac{1}{10}u^{10} - \frac{2}{9}u^9 + \frac{1}{8}u^8 + c\\ &= \frac{1}{10}(t+1)^{10} - \frac{2}{9}(t+1)^9 + \frac{1}{8}(t+1)^8 + c.\end{aligned}$$

Here we used the substitution $u = t + 1$. Then $du = dx$ and $t = u - 1$.

We may have to use a substitution and a trigonometric identity to solve

[3]We do not attach any particular meaning to the symbols dx and du in their own right. The equation $du = g'(x)dx$ helps us to write down what happens when we perform the substitution as in the first equality in (3.15). Thought of as infinitesimals or differentials, these symbols have a meaning, but this is beyond the scope of these notes.

an integration problem:

$$\begin{aligned}
\int 2x\sin^2(x^2+5)\ dx &= \int \sin^2 u\ du \\
&= \frac{1}{2}\int [1-\cos 2u]\ du \\
&= \frac{1}{2}\left[u-\frac{1}{2}\sin 2u\right]+c \\
&= \frac{1}{2}\left[(x^2+5)-\frac{1}{2}\sin[2(x^2+5)]\right]+c.
\end{aligned}$$

We used the substitution $u=x^2+5$, so that $du=2xdx$ and the identity $\sin^2\alpha=[1-\cos 2\alpha]/2$.

Find the substitution which we used in the following computation, and check the details:

$$\begin{aligned}
\int \sec^2 x\tan x\ dx &= \int \sec x\cdot\sec x\tan x\ dx \\
&= \int u\ du \\
&= \frac{1}{2}u^2+c \\
&= \frac{1}{2}\sec^2 x+c.
\end{aligned}$$

Sometimes we have to apply the method of substitution twice, or more often, to work out an integral. Here is an example.

$$\begin{aligned}
\int (x^2+1)\sin^3(x^3+3x-2)\cos(x^3+3x-2)\ dx &= \frac{1}{3}\int \sin^3 u\cos u\ du \\
&= \frac{1}{3}\int v^3\ dv \\
&= \frac{1}{12}v^4+c \\
&= \frac{1}{12}\sin^4 u+c \\
&= \frac{\sin^4(x^3+3x-2)}{12}+c
\end{aligned}$$

In the computation we used the substitution $u=x^3+3x-2$. Then $du=3(x^2+1)\ dx$. In a second substitution we set $v=\sin u$. Then $dv=\cos u\ du$.

Here are two examples, which are important in the context of integrating rational functions. In the first example we assume that $a \neq 0$, and we use the subsitution $x = au$. The $dx = adu$.

$$\begin{aligned} \int \frac{dx}{x^2+a^2}\ dx &= \int \frac{adu}{a^2u^2+a^2} \\ &= \frac{1}{a}\int \frac{du}{u^2+1} \\ &= \frac{1}{a}\arctan(u)+c \\ &= \frac{1}{a}\arctan\left(\frac{x}{a}\right)+c \end{aligned}$$

Adding another idea, we calculate

$$\int \frac{dx}{x^2+2x+5} = \int \frac{dx}{(x+1)^2+4} = \int \frac{du}{u^2+4} = \frac{1}{2}\arctan\left(\frac{x+1}{2}\right)+c.$$

We used the substitution $u = x+1$, and then we proceeded as in the previous example.

3.9.1 Substitution and Definite Integrals

Let us now explore how substitution is used to calculate definite integrals. Assuming as before that f and g' are continuous on the interval $[a, b]$, we have

$$\int_a^b f(g(x))g'(x)\ dx = \int_{g(a)}^{g(b)} f(u)\ du. \tag{3.16}$$

To see this, observe that f has an anti-derivative, which we again denote by F. Then

$$\int_a^b f(g(x))g'(x)\ dx = F(g(x))\Big|_a^b = F(u)\Big|_{g(a)}^{g(b)} = \int_{g(a)}^{g(b)} f(u)\ du.$$

The first identity is obtained as a combination of the Fundamental Theorem of Calculus and (3.14). The second one is obvious, and the third one is another application of the Fundamental Theorem of Calculus.

Let us apply this formula in a few examples.

$$\int_0^1 (x^2-1)(x^3-3x+5)^3\ dx = \frac{1}{3}\int_5^3 u^3\ du = \frac{1}{12}u^4\Big|_5^3 = -\frac{136}{3}.$$

We used the substitution $u = x^3 - 3x + 5$. Then $du = (3x^2 - 3)\ dx$, and $\frac{1}{3}du = (x^2 - 1)\ dx$. To obtain the limits for the integral we calculate $u(0) = 5$ and $u(1) = 3$.

Another example is

$$\int_0^{\pi/4} \cos^2 x \sin x\ dx = -\int_1^{\sqrt{2}/2} u^2\ du = -\frac{1}{3}u^3\Big|_1^{\sqrt{2}/2} = -\frac{1}{3}\left[1 - \frac{\sqrt{2}}{4}\right].$$

We use the substitution $u = \cos x$. Then $-du = \sin x\ dx$. If $x = 0$, then $u = 1$, and if $x = \pi/4$, then $u = \sqrt{2}/2$.

Incorporating one of our previous techniques, we calculate

$$\int_0^2 x(x+1)^6\ dx = \int_1^3 (u-1)u^6\ du = \int_1^3 u^7 - u^6\ du = \frac{3554}{7}.$$

We use the substitution $u = x + 1$. Then $du = dx$ and $x = u - 1$. If $x = 0$, then $u = 1$, and if $x = 2$, the $u = 3$.

Similarly,

$$\int_0^{\sqrt{8}} x^3\sqrt{x^2+1}\ dx = \frac{1}{2}\int_1^9 (u-1)\sqrt{u}\ du = \frac{1}{2}\int_1^9 \left(u^{3/2} - u^{1/2}\right)\ du = \frac{596}{15}.$$

We use the substitution $u = x^2 + 1$. Then $\frac{1}{2}du = x\ dx$ and $x^2 = u - 1$. For the limits we calculate, if $x = 0$, then $u = 1$, and if $x = \sqrt{8}$, then $u = 9$.

Finally,

$$\int_0^1 \sqrt{1 - x^2}\ dx = \int_0^{\pi/2} \sqrt{1 - \sin^2 u}\cos u\ du = \int_0^{\pi/2} \cos^2 u\ du = \frac{\pi}{4}.$$

We use the substitution $x = \sin u$. Then $dx = \cos u\ du$. If $x = 0$, then $u = 0$, and if $x = 1$, then $u = \pi/2$. For our given values of x, there are other possible values for u, but they will lead to the same results.

Remark 12. The graph of $f(x) = \sqrt{1 - x^2}$ is the northern part of a circle. Using $x \in [0, 1]$ means that we calculated the area under this graph in the first quadrant, i.e., the area of one forth of the disk of radius 1. You were told long time ago in school, that the area of this unit disk is π, so that the result of the calculation is hardly surprising.

There is a more serious matter. Is the example genuine, or did we assume the answer previously? By definition, π is the ratio of the circumference of a circle by its diameter. In our calculation of the derivative of the sine and cosine functions we used the estimate that $|\sin h - h| \leq h^2/2$. When we showed this, we used that $|h| \le |\tan h|$ for $h \in [-\pi/4, \pi/4]$. A typical proof of the latter inequality starts out by first showing that the area of the unit disk is π. This means, we assumed the result in the example, we did not derive it.

Exercise 48. Find the following integrals:

(a) $\displaystyle\int \frac{dx}{\sqrt{2x+1}}$

(b) $\displaystyle\int \frac{t}{(4t^2+9)^2}\, dt$

(c) $\displaystyle\int t(1+t^2)^3\, dt$

(d) $\displaystyle\int \frac{2s}{\sqrt[3]{6-5s^2}}\, ds$

(e) $\displaystyle\int \frac{b^3x^3}{\sqrt{1-a^4x^4}}\, dx$

(f) $\displaystyle\int_0^{\pi} x\cos x^2\, dx$

(g) $\displaystyle\int x^2\sqrt{x+1}\, dx$

(h) $\displaystyle\int \frac{x+3}{\sqrt{x+1}}\, dx$

(i) $\displaystyle\int \sin^2(3x)\, dx$

(j) $\displaystyle\int_0^{\pi/2} \cos^2 x\, dx$

(k) $\displaystyle\int_{\pi/6}^{\pi/4} \sec(2x)\tan(2x)\, dx$

(l) $\displaystyle\int_0^{1/2} \frac{dx}{4+x^2}$

(m) $\displaystyle\int \frac{\sec^2 x}{\sqrt{1+\tan x}}\, dx$

(n) $\displaystyle\int \sqrt{1+\sin x}\cos x\, dx$

(o) $\displaystyle\int_0^r \sqrt{r^2-x^2}\, dx$

3.10 Areas between Graphs

Previously we related the integral to areas of a region under a graph. This idea can be generalized to the discussion of areas of regions between two graphs. Let us look at an example.

Example 3.28. Calculate the area of the region between the graphs of the functions $f(x) = x^2$ and $g(x) = \sqrt{1-x^2}$.

Solution To get a better understanding, we draw the two graphs, see Figure 3.13. Now you see the region between the two graphs whose area we want to calculate. We call the region Ω.

The graphs intersect in two points. To find their x-coordinates, we solve the equation

$$f(x) = x^2 = g(x) = \sqrt{1-x^2}.$$

After squaring the equation and solving it for x^2, we find $x^2 = \frac{-1\pm\sqrt{5}}{2}$. Only the $+$ sign occurs as $x^2 \ge 0$. Taking the square root, we find the x-coordinates of the points where the curves intersect:

$$A = -\sqrt{\frac{-1+\sqrt{5}}{2}} \quad \text{and} \quad B = \sqrt{\frac{-1+\sqrt{5}}{2}}.$$

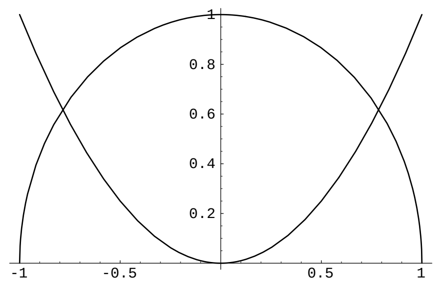

Figure 3.13: Region between two graphs

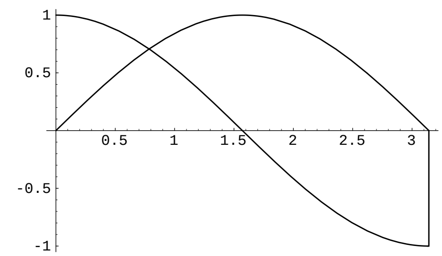

Figure 3.14: Region between two graphs

To get the area of the region under the graph of $f(x)$ and $g(x)$ over the interval $[A, B]$ we can calculate the appropriate integrals. To get the area of the region Ω between the graphs, we take the area of the region under the graph of $g(x)$ and subtract the area of the region under the graph of $f(x)$. Concretely:

$$\text{Area}(\Omega) = \int_A^B g(x)\ dx - \int_A^B f(x)\ dx = \int_A^B (g(x) - f(x))\ dx \approx 1.06651.$$

The numerical value was obtained by computer. You are invited to work out the integral with the help of the Fundamental Theorem of Calculus to verify the result. $\Diamond$

Some problems are a bit more subtle.

Example 3.29. Find the area of the region between the graphs of the functions $f(x) = \cos x$ and $g(x) = \sin x$ for x between 0 and π.

Solution: The region Ω between the graphs is shown in Figure 3.14. The region breaks up into two pieces, the region Ω_1 over the interval $[0, \pi/4]$ on which $f(x) \geq g(x)$, and the region Ω_2 over the interval $[\pi/4, \pi]$ where $g(x) \geq f(x)$. We calculate the areas of the regions Ω_1 and Ω_2 separately.

In each case, we proceed as in the previous example:

$$\text{Area}(\Omega_1) = \int_0^{\pi/4} (\cos x - \sin x)\ dx = (\sin x + \cos x)\Big|_0^{\pi/4} = \sqrt{2} - 1$$
$$\text{Area}(\Omega_2) = \int_{\pi/4}^{\pi} (\sin x - \cos x)\ dx = -(\sin x + \cos x)\Big|_{\pi/4}^{\pi} = 1 + \sqrt{2}.$$

In summary we find:

$$\text{Area}(\Omega) = \text{Area}(\Omega_1) + \text{Area}(\Omega_2) = 2\sqrt{2}.$$

An additional remark may be in place. When we compared integrals and areas, we had to take into account where the function is non-negative, resp., non-positive. Here we did not. We took care of this aspect by breaking up the interval into the part where $f(x) \geq g(x)$ and the part where $g(x) \geq f(x)$. $\diamondsuit$

Our general definition for the area between two graphs is as follows.

Definition 3.30. *Suppose $f(x)$ and $g(x)$ are integrable functions over an interval $[a, b]$. Let Ω be the region between the graphs of $f(x)$ and $g(x)$ for x between a and b. The area of Ω is*

$$\text{Area}(\Omega) = \int_a^b |f(x) - g(x)|\ dx.$$

This definition generalizes Definition 3.7 on page 114. The definition is also consistent with the intuitive idea of the area of a region, and it incorporates and generalizes Proposition 3.23 on page 121. Taking the absolute value of the difference of $f(x)$ and $g(x)$ allows us avoid the question where $f(x) \geq g(x)$ and where $g(x) \geq f(x)$. Typically this problem gets addressed when the integral is calculated. In some problems a and b are explicitly given, in others you have to determine them from context. In all cases it is good to graph the functions before calculating the area of the region between them. Having the correct picture in mind helps you to avoid mistakes.

Exercise 49. Sketch and find the area of the region bounded by the curves:

(a) $y = x^2$ and $y = x^3$.

(b) $y = 8 - x^2$ and $y = x^2$

(c) $y = x^2$ and $y = 3x + 5$.

(d) $y = \sin x$ and $y = \pi x - x^2$.

(e) $y = \sin x$ and $y = 2 \sin x \cos x$ for x between 0 and π.

3.11 Numerical Integration

The Fundamental Theorem of Calculus provided us with a highly efficient method for calculating definite integrals. Still, for some functions we have no good expression for its anti-derivative. In such cases we may have to rely on numerical methods for integrating. Let us take such a function, and show some methods for finding an approximate value for the integral.

We describe different ways to find, by numerical means, approximate values for the integral of a function $f(x)$ over the interval $[a, b]$:

$$\int_a^b f(x)\ dx.$$

In all of the different approaches we partition the interval into smaller ones:

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

Left and Right Endpoint Method: In the left endpoint method we find the value of the function at each left endpoint of the intervals of the partition. We multiply it with the length of the associated interval, and then add up the terms. Explicitly, we calculate

$$I_L = f(x_0)(x_1 - x_0) + f(x_1)(x_2 - x_1) + \cdots + f(x_{n-1})(x_n - x_{n-1}). \tag{3.17}$$

In the right endpoint method we proceed as we did on the left endpoint method, only we use the value of the function at the right endpoint instead of the left endpoint:

$$I_R = f(x_1)(x_1 - x_0) + f(x_2)(x_2 - x_1) + \cdots + f(x_n)(x_n - x_{n-1}). \tag{3.18}$$

Both expressions provide us with specific examples of Riemann sums.

Example 3.31. Use the left and right endpoint method to find approximate values for

$$\int_0^2 e^{-x^2}\ dx.$$

Solution: Set $f(x) = e^{-x^2}$ and choose the partition:

$$x_0 = 0 < x_1 = \frac{1}{2} < x_2 = 1 < x_3 = \frac{3}{2} < x_4 = 2.$$

Then $x_k - x_{k-1} = 1/2$ for $k = 1, 2, 3$, and 4. Formula (3.17) for I_L specializes to

$$I_L = \frac{f(0) + f(1/2) + f(1) + f(3/2)}{2} \approx 1.126039724.$$

Formula (3.18) for I_R specializes to

$$I_R = \frac{f(1/2) + f(1) + f(3/2) + f(2)}{2} \approx .6351975438.$$

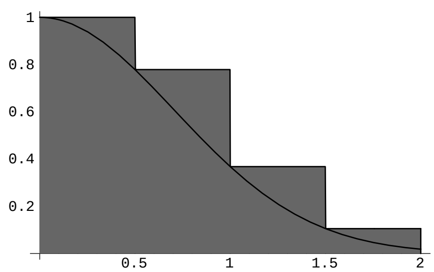

Figure 3.15: Use left end points

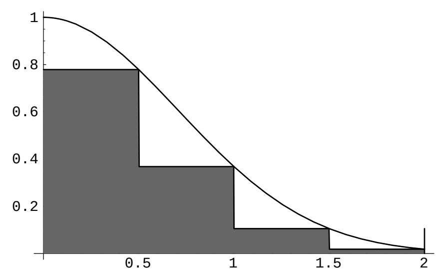

Figure 3.16: Use right end points

Apparently, I_L and I_R are calculated by combining the areas of certain rectangles. In our case the values of $f(x)$ are all positive and all of the rectangles are above the x axis, so the areas of the rectangles are all added. Note also, that our specific function $f(x)$ is decreasing on the interval $[0, 2]$, so that I_L is an upper sum for the function $f(x)$ over the interval $[0, 2]$, and I_R is a lower sum. In this sense, we have

$$I_R = .6351975438 \leq \int_0^2 e^{-x^2}\ dx \leq I_L = 1.126039724.$$

The function and the rectangles whose areas are added to give us I_L and I_R are shown in Figure 3.15 and Figure 3.16. $\Diamond$

Midpoint and Trapezoid Method: We may try and improve on the endpoint methods. In the midpoint methods, we use the value of the function at the midpoints of the intervals of the partition. That should be less bias.

We use the same partition and notation as above. Then the formula for the midpoint method is:

$$I_M = f\left(\frac{x_0 + x_1}{2}\right)(x_1 - x_0) + \cdots + f\left(\frac{x_n + x_{n-1}}{2}\right)(x_n - x_{n-1}). \tag{3.19}$$

In the trapezoid method we do not take the function at the average (i.e. midpoint) of the end points of the intervals in the partition, but we average the values of the function at the end points. Specifically, the formula is

$$I_T = \frac{f(x_0) + f(x_1)}{2}(x_1 - x_0) + \cdots + \frac{f(x_{n-1}) + f(x_n)}{2}(x_n - x_{n-1}). \tag{3.20}$$

It is quite easy to see that

$$I_T = \frac{I_L + I_R}{2}. \tag{3.21}$$

Let us explain the reference to the word trapezoid. For simplicity, suppose that $f(x)$ is non-negative on the interval $[a, b]$. Consider the trapezoid of width $(x_1 - x_0)$ which has height $f(x_0)$ at its left and $f(x_1)$ at its right edge. The area of this trapezoid is $\frac{f(x_0)+f(x_1)}{2}(x_1 - x_0)$. This is the first summand in the formula for I_T, see (3.20). We have such a trapezoid over each of the intervals in the partition, and their areas are added to give I_T.

Expressed differently, we can draw a secant line through the points $(x_0, f(x_0))$ and $(x_1, f(x_1))$. This gives us the graph of a function $T(x)$ over the interval $[x_0, x_1]$. Over the interval $[x_1, x_2]$ the graph of $T(x)$ is the secant line through the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$. Proceeding in the fashion, we use appropriate secant lines above all of the intervals in the partition to define the function $T(x)$ over the entire interval $[a, b]$. Then

$$I_T = \int_a^b T(x)\, dx.$$

This integral is easily computed by the formula in (3.20).

Example 3.32. Use the midpoint and trapezoid method to find approximate values for

$$\int_0^2 e^{-x^2}\, dx.$$

Solution: We use the same partition of $[0, 2]$ as in Example 3.31. The formula for I_M (see (3.19)) specializes to

$$I_M = \frac{f(.25) + f(.75) + f(1.25) + f(1.75)}{2} \approx .8827889485.$$

As for the endpoint methods, I_M is the combined area of certain rectangles. Their heights are the values $f(x_i)$ at the midpoints of the intervals of the partition. Their width are the lengths of the intervals of the partition. You see the rectangles for this calculation in Figure 3.17.

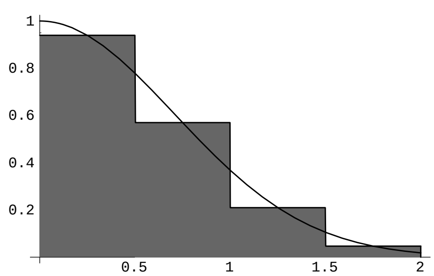

Figure 3.17: Use midpoints

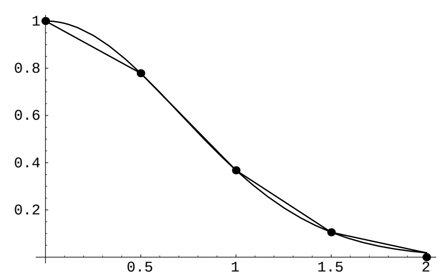

Figure 3.18: Trapezoid Method

Based on our previous calculations and Formula (3.21) we find

$$I_T = \frac{I_L + I_R}{2} \approx .8806186341.$$

We illustrated this calculation in Figure 3.18. There you see the function $f(x) = e^{-x^2}$ and five dots on the graph. The dots are connected by straight line segments. These line segments form the graph of a function $T(x)$, and I_T is the area of the region under this graph. So

$$I_T = \int_0^2 T(x)\, dx. \quad \Diamond$$

Simpson's Method: In Simpson's method we combine the endpoint and midpoint methods in a weighted fashion. Again, we use the same notation for the function and the partition as above. The specific formula for an approximate value of the integral of $f(x)$ over $[a, b]$ is

$$\begin{aligned} I_S =& \frac{1}{6}\left[f(x_0) + 4f\left(\frac{x_0 + x_1}{2}\right) + f(x_1)\right](x_1 - x_0) + \cdots \\ &+ \frac{1}{6}\left[f(x_{n-1}) + 4f\left(\frac{x_{n-1} + x_n}{2}\right) + f(x_n)\right](x_n - x_{n-1}) \end{aligned} \tag{3.22}$$

It is quite easy to see that

$$I_S = \frac{I_L + 4I_M + I_R}{6} = \frac{I_T + 2I_M}{3}.$$

Let us explain the background to Simpson's method. We define a function $P(x)$ over the interval $[a, b]$ by defining a degree 2 polynomial on each of the intervals of the partition. The polynomial over the interval $[x_{k-1}, x_k]$ is chosen so that it agrees with $f(x)$ at the end points and at the midpoint of this interval. Simpson's method is a refinement of the Trapezoid method. In one method we use two points on the graph and connect them by a straight line segment. In the other one we use three points on the graph and construct a parabola through them. With some work one can show that

$$I_S = \int_a^b P(x)\ dx.$$

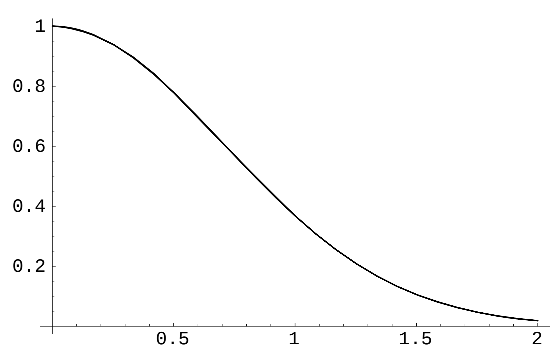

Figure 3.19: Simpson's Method

Example 3.33. Use Simpson's method to find an approximate value for

$$\int_0^2 e^{-x^2}\ dx.$$

Solution: We use the same partition of $[0, 2]$ as in Example 3.31. The formula for I_S (see the special case of (3.22)) specializes to

$$I_S = \frac{I_L + 4I_M + I_R}{6} \approx .88206555104,$$

where I_L, I_M and I_R are as above.

You see the method illustrated in Figure 3.19. There you see the graphs of two functions, the function $f(x) = e^{-x^2}$ and the function $P(x)$ from the discussion of Simpson's method. Only the thickness of the line suggests that there are two graphs of almost identical functions. $\diamondsuit$

Example 3.34. Compare the accuracy of the various approximate values of

$$\int_0^2 e^{-x^2}\, dx.$$

Solution: We compare the approximate values for the integral obtained by the different formulas. We partition the interval $[0, 2]$ into n intervals of the same length, and vary n. We tabulate the results. They should be compared with an approximate value for the integral of

0.882081390762421.

	$n = 1$	$n = 10$	$n = 100$	$n = 1000$
I_L	2.0000000	0.9800072469	0.891895792451	0.883063050702697
I_R	0.0366313	0.7836703747	0.872262105229	0.881095681980474
I_M	0.7357589	0.8822020700	0.882082611663	0.882081402972833
I_T	1.0183156	0.8818388108	0.882078948840	0.882081366341586
I_S	0.8299445	0.8820809836	0.882081390722	0.882081390762417

Table 3.1: Approximate Values of the Integral

Simpson's method is more accurate than the other ones. E.g., Simpson's method with $n = 4$ gives a result which is better than the left and right endpoint method with $n = 1000$. Even if you use the midpoint and trapezoid method with $n = 1000$, then the result is far less accurate that Simpson's method with $n = 100$. $\diamondsuit$

Remark 13. It is important that we keep the number n of intervals into which we partition $[a, b]$ small. It does not only keep the number of overall computations small. In each computational step we expect to make a round-off error, and these may add up. The fewer computations we make, the smaller the cummulative round-off error will be.

Exercise 50. Proceed as in Example 3.34 and compare the different methods applied to the calculation of

$$\int_0^{\pi/2} \sin x \; dx = 1.$$

3.12 Applications of the Integral

In Definition 3.7 and Proposition 3.23 we related definite integrals to areas. Based on the context, this can have a more concrete meaning. Consider a function $f(t)$ on an interval $[a, b]$ and the integral

$$I = \int_a^b f(t) \; dt.$$

If $f(t)$ stands for the rate at which a drug is absorbed, then I is the total amount of the drug which has been absorbed in the time interval $[a, b]$. If $f(t)$ stands for the speed with which you travel, then I stands for the total distance which you traveled during the time interval $[a, b]$. You are invited to come up with more interpretations. In addition, the following definition expresses the common notion of the average value of a function.

Definition 3.35. *Suppose that $f(t)$ is an integrable function over the interval $[a, b]$. Then the quantity*

$$f_{av} := \frac{1}{b-a} \int_a^b f(t) \; dt$$

is called the average value *of $f(t)$ over the interval $[a, b]$.*

For example, the average value of the sine function $f(x) = \sin x$ over the interval $[0, \pi]$ is $2/\pi$.

Let us explore the different aspects of integration in an example.

Example 3.36. The river Little Brook flows into a reservoir, referred to as Beaver Pond by the locals. The amount of water carried by the river depends on the season. As a function of time, it is

$$g(t) = 2 + \sin\left(\frac{\pi t}{180}\right).$$

We measure time in days, and $t = 0$ corresponds to New Year. The units of $g(t)$ are millions of liter of water per day. Water is released from Beaver Pond at a constant rate of 2 million liters per day. At the beginning of the year, there are 200 million liters of water in the reservoir.

(a) How many liter of water are in Beaver Pond by the end of April?

(b) Suppose $F(t)$ tells how much water there is in the reservoir on day t of the year. Find $F(t)$.

(c) At which rate does the amount of water in the reservoir change at the beginning of September?

(d) On which days will there be 250 million liters of water in Beaver Pond?

(e) At which amount of water will the reservoir crest?

(f) On the average, by how much has the amount of water in Beaver Pond increased per day during the first three months of the year?

Solution: Water enters and leaves the pond. The net rate entering is

$$f(t) = g(t) - 2 = \sin\left(\frac{\pi t}{180}\right) \text{ millions of liters per day.}$$

We obtain the total change of the amount of water in the reservoir by integrating $f(t)$. Set

$$A(T) = \int_0^T f(t)\, dt.$$

On the T-th day of the year, the total amount of water in Beaver Pond is

$$F(T) = 200 + \int_0^T f(t)\, dt = 200 + \frac{180}{\pi}\left[1 - \cos\left(\frac{\pi T}{180}\right)\right] \text{ millions of liters.}$$

This answers (b). By the end of April, after 120 days, there are

$$F(120) = 200 + \frac{180}{\pi}\left[1 - \cos\frac{2\pi}{3}\right] \approx 238.2$$

millions of liters of water in the pond. This answers (a).

The rate at which the amount of water in the pond changes is $F'(t) = f(t)$. At the beginning of September, after 240 days, the rate of change is

$f(240) \approx -.866$. The pond is losing water at a rate of 866,000 liters per day.

To answer (d), we like to know for which T we have $F(T) = 250$. We solve the equation for T:

$$250 = 200 + \frac{180}{\pi}\left[1 - \cos\left(\frac{\pi T}{180}\right)\right] \quad \text{or} \quad \cos\left(\frac{\pi T}{180}\right) = 1 - \frac{5\pi}{18}.$$

We apply the function arccos to both sides of the last equation and find

$$T = \frac{180}{\pi}\arccos\left(1 - \frac{5\pi}{18}\right) \approx 88, \text{ or } 272.$$

On the 88-th and 272-nd day of the year there will be 250 millions of liters of water in the reservoir.

To find at which amount the reservoir crests, we have to find the maximum value of $F(t)$. This occurs apparently when $\cos(\pi t/180) = -1$ or $t = 180$. The pond crests at mid-year, and then the amount of water in it is about 314.6 millions of liters of water. This answers (e).

After three months or 90 days there are about 257.3 millions of liters of water in Beaver Pond. Within this time, the amount of water has increased by 57.3 millions of liters. On the average, the amount of water in the reservoir increased by about 640,000 liters per day. ♢

Exercise 51. A pain reliever has been formulated such that it is absorbed at a rate of $600 \sin(\pi t)$ (mg/hr) by the body. Here t measures time in hours, $t = 0$ at the time you take the medication, and the absorption process is complete at time $t = 1$.

(a) What is the total amount of the drug which is absorbed?

(b) Find a function $F(t)$, such that $F(t)$ tells how much medication has been absorbed at time t.

(c) A total of 150 mg of the medication has to be absorbed before the drug is effective. How long does it take until this threshold is reached?

3.13 The Exponential and Logarithm Functions

In Section 1.10 we introduced the *exponential function* $\exp(x) = e^x$ and the *natural logarithm function* $\ln x$. At the time we only stated that they exist because we did not have the tools to properly define them. We will now fill in the details. Many of the routine calculations are formulated as exercises.

Definition 3.37. *Let $x \in (0, \infty)$. The* natural logarithm *of x is defined as*

$$\ln x = \int_1^x \frac{dt}{t}. \tag{3.23}$$

Theorem 3.38. *The natural logarithm function is differentiable on its entire domain $(0, \infty)$, its derivative is*

$$\ln' x = \frac{1}{x},$$

and $\ln x$ is increasing on $(0, \infty)$.

Proof. The function $1/x$ is defined and continuous on $(0, \infty)$. According to Theorem 3.10 this means that $\ln x$ is defined for all x in $(0, \infty)$. Theorem 3.26 tells us that $\ln' x = 1/x$. According to Theorem 2.11, the function is increasing because its derivative $\ln' x > 0$ for all $x > 0$. □

Let us also verify one of the central equations for calculating with logarithms, the third rule in Theorem 1.34.

Proposition 3.39. *For any $x,\ y > 0$,*

$$\ln(xy) = \ln x + \ln y. \tag{3.24}$$

Proof. We need a short calculation. Here x and y are fixed positive numbers. We use the substitution $u = \frac{t}{x}$, so that $du = \frac{1}{x}\,dt$. For the adjustment of the limits of integration, observe that $t/x = u = 1$ when $t = x$, and that $t/x = u = y$ when $t = xy$. Then

$$\int_x^{xy} \frac{dt}{t} = \int_x^{xy} \frac{1}{(t/x)} \frac{1}{x}\, dt = \int_1^y \frac{du}{u} = \ln y.$$

Using this calculation we deduce that

$$\ln(xy) = \int_1^{xy} \frac{dt}{t} = \int_1^x \frac{dt}{t} + \int_x^{xy} \frac{dt}{t} = \ln x + \ln y.$$

This is exactly our claim. □

Exercise 52. Show:

(1) $\ln 1 = 0$.

(2) $\ln(1/y) = -\ln y$ for all $y > 0$.

(3) $\ln(x/y) = \ln x - \ln y$ for all x, $y > 0$.

Exercise 53. Show that $\ln 4 > 1$. Hint: Using the partition

$$1 = x_0 < 2 = x_1 < 3 = x_2 < 4 = x_3,$$

find a lower sum S_l for the function $1/t$ over the interval $[1, 4]$ so that $S_l > 1$.

We can now define the Euler number:

Definition 3.40. *The number* Euler number e *is the unique number such that*

$$\ln e = 1 \quad \textit{or, equivalently,} \quad \int_1^e \frac{dt}{t} = 1.$$

For this definition to make sense, we have to show that there is a number e which has the property used in the definition. To see this, observe that $\ln 1 = 0 < 1 < \ln 4$. Because $\ln x$ is differentiable, it follows from the Intermediate Value Theorem (see Theorem 1.16) that there is a number e for which $\ln e = 1$. It also follows that $1 < e < 4$.

Proposition 3.41. *For every real number x there exists exactly one positive number y, such that*

$$\ln y = x \tag{3.25}$$

Proof. Observe that $\ln(e^n) = n$ and $\ln(1/e^n) = -n$ for all natural numbers n. So all integers (whole numbers) are values of the natural logarithm function. Every real number x lies between two integers. According to the Intermediate Value Theorem, every real number is a value of the function $\ln y$. We saw that $\ln y$ is an increasing function. This means that, for any given x, the equation $\ln y = x$ has at most one solution. Taken together it means that it has a unique solution. ☐

Exercise 54. Show that

$$\ln(a^r) = r \ln a$$

for all positive numbers a and all rational numbers r, i.e., numbers of the form $r = p/q$ where p and q are integers and $q \neq 0$.

In summary, we have seen that

Corollary 3.42. *The natural logarithm function $\ln x$ is a differentiable, increasing function with domain $(0, \infty)$ and range $(-\infty, \infty)$, and $\ln' x = 1/x$.*

We are now ready to define the *exponential function.*

Definition 3.43. *Given any real number x, we define* $\exp(x)$ *to be the unique number for which*

$$\ln(\exp(x)) = x, \tag{3.26}$$

i.e., $y = \exp(x)$ *is the unique solution of the equation* $\ln(y) = x$*. This assignment (mapping x to* $\exp(x)$*) defines a function, called the* exponential function*, with domain* $(-\infty, \infty)$ *and range* $(0, \infty)$*.*

Exercise 55. Show that the exponential function exp and the natural logarithm function ln are inverses of each other. In addition to the equation in (3.26), you need to show that

$$\exp(\ln(y)) = y \tag{3.27}$$

for all $y \in (0, \infty)$.

Summarizing this discussion, and adding some observations which we have made elsewhere, we have:

Proposition 3.44. *The exponential function* $\exp(x)$ *is a differentiable, increasing function with domain* $(-\infty, \infty)$ *and range* $(0, \infty)$*, and the exponential function is its own derivative, i.e.,* $\exp'(x) = \exp(x)$*.*

Exercise 56. Show for all real numbers x and y that:

(1) $\exp(0) = 1$

(2) $\exp(1) = e$

(3) $\exp(x)\exp(y) = \exp(x + y)$

(4) $1/\exp(y) = \exp(-y)$

(5) $\exp(x)/\exp(y) = \exp(x - y)$.

Hint: Use the results of Exercise 52, the definition of e in Definition 3.40, and that the exponential and logarithm functions are inverses of each other.

Exercise 57. Show that $\exp(r) = e^r$ for all rational numbers r. Hint: Use Exercise 54 and that the exponential and logarithm functions are inverses of each other.

The expression e^r makes sense only if r is a rational number. If $r = p/q$ then we raise e to the r-th power and take the q-th root of the result. For an arbitrary real number we set

$$e^x = \exp(x). \tag{3.28}$$

This is consistent with the meaning of the expression for rational exponents due to Exercise 57, and it defines what we mean by raising e to any real power.

3.13.1 Other Bases

So far we discussed the natural logarithm function and the exponential function with base e. We now expand the discussion to other bases.

Definition 3.45. *Let a be a positive number, $a \neq 1$. Set*

$$\log_a x = \frac{\ln x}{\ln a} \quad \textit{and} \quad \exp_a(x) = \exp(x \ln a). \tag{3.29}$$

We call $\log_a(x)$ *the* logarithm function with base a *and* $\exp_a(x)$ *the* exponential function with base a. *For the function* $\log_a$ *we use the domain* $(0, \infty)$ *and range* $(-\infty, \infty)$. *For the exponential function* $\exp_a$ *we use the domain* $(-\infty, \infty)$ *and range* $(0, \infty)$.

Exercise 58. Show

(1) $\ln a > 0$ if $a > 1$ and $\ln a < 0$ if $0 < a < 1$.

(2) $\log_a(x)$ and $\exp_a(x)$ are differentiable functions.

(3) $\log_a(x)$ and $\exp_a(x)$ are increasing functions if $a > 1$.

(4) $\log_a(x)$ and $\exp_a(x)$ are decreasing functions if $0 < a < 1$.

Exercise 59. Suppose $a > 0$ and $a \neq 1$. Show that

(a) $\exp_a(\log_a(y)) = y$ for all $y > 0$.

(b) $\log_a(\exp_a(x)) = x$ for all real numbers x.

Taken together, the specifications for the domains and ranges for the functions $\exp_a$ and $\log_a$ and the results from Exercise 59 tell us that

Corollary 3.46. *Suppose $a > 0$ and $a \neq 1$. The functions* $\exp_a$ *and* $\log_a$ *are inverses of each other.*

Exercise 60. Suppose $a > 0$ and $a \neq 1$. Show the *laws of logarithms*:

(a) $\log_a 1 = 0$ and $log_a a = 1$.

(b) $\log_a(xy) = \log_a x + \log_a y$ for all x, $y > 0$.

(c) $\log_a(1/y) = -\log_a y$ for all $y > 0$.

(d) $\log_a(x/y) = \log_a x - \log_a y$ for all x, $y > 0$.

Exercise 61. Suppose $a > 0$ and $a \neq 1$. Show the *exponential laws*:

(1) $\exp_a(0) = 1$ and $\exp_a(1) = a$

(2) $\exp_a(x)\exp_a(y) = \exp_a(x + y)$

(3) $1/\exp_a(y) = \exp_a(-y)$

(4) $\exp_a(x)/\exp_a(y) = \exp_a(x - y)$.

Exercise 62. Suppose $a > 0$, $a \neq 1$, and r is a rational number. Show

$$\log_a(a^r) = r \quad \text{and} \quad \exp_a(r) = a^r.$$

We rephrase a convention which we made previously for e. Suppose $a > 0$ and $a \neq 1$. The expression a^r makes sense if r is a rational number. If $r = p/q$ then we raise a to the r-th power and take the q-th root of the result. For an arbitrary real number we set

$$a^x = \exp_a(x). \tag{3.30}$$

This is consistent with the meaning of the expression for rational exponents due to Exercise 62, and it defines what we mean by raising a to any real power. Equation 3.30 specializes to the one in Equation 3.28 if we set $a = e$. It is also a standard convention to set

$$1^x = 1 \quad \text{and} \quad 0^x = 0$$

for any real number x. Typically 0^0 is set 1.

We can now state an equation which is typically considered to be one of the laws of logarithms:

Exercise 63. Suppose $a > 0$, $a \neq 1$, $x > 0$, and z is any real number. Then

$$\log_a(x^z) = z\log_a(x).$$

We are now ready to fill in the details for one of the major statements which we made in Section 1.10. We are ready to prove

Theorem 3.47. *Let a be a positive number, $a \neq 1$. There exists exactly one monotonic function, called the exponential function with base a and denoted by* $\exp_a(x)$, *which is defined for all real numbers x such that* $\exp_a(x) = a^x$ *whenever x is a rational number.*

Proof. In this section we constructed the function $\exp_a(x)$, and this function has all of the properties called for in the theorem. That settles the existence statement. We have to show the uniqueness statement, i.e., there in only one such function.

Suppose $f(x)$ is any monotonic function and $f(r) = a^r = \exp_a(r)$ for all rational numbers r. We have to show that $f(x) = \exp_a(x)$ for all real numbers x. We leave the verification of this assertion to the reader. Here one uses that $f(x)$ and $\exp_a(x)$ are monotonic, and that $\exp_a(x)$ is continuous. □

Chapter 4

Trigonometric Functions

In this section we discuss the radian measure of angles and introduce the trigonometric functions. These are the functions sine, cosine, tangent, et. al. We collect some formulas relating these functions.

Arc Length and Radian Measure of Angles: Consider the unit circle (a circle with radius 1) centered at the origin in the Cartesian plane. It is shown in Figure 4.1. We take a practical approach to measuring the length of an arc on this circle. We imagine that we can straighten it out, and measure how long it is. It requires some work to introduce the idea of the length of a curve in a mathematically rigorous fashion.

Definition 4.1. *The number π is the ratio between the circumference of a circle and its diameter.*

This definition goes back to the Greeks. Stated differently it says, that the circumference of a circle of radius r is $2\pi r$. Observe that the ratio referred to in the definition does not depend on the radius of the circle.

Consider an angle α between the positive x-axis and a ray which originates at the origin of the coordinate system and intersects the unit circle in the point p. We like to find the radian measure of the angle α. Consider an arc on the unit circle which starts out at the point $(1, 0)$ and ends at p, and suppose its length is s. Then

$$\alpha = \pm s \text{ (radians)}. \tag{4.1}$$

The $+$ sign is used if the arc goes counter clockwise around the circle. The $-$ sign is used if it proceeds clockwise. We may also consider arcs which wrap around the circle several times before they end at p. In this sense, the radian measure of the angle α is not unique, but any two radian measures of the angle differ by an integer multiple of 2π.

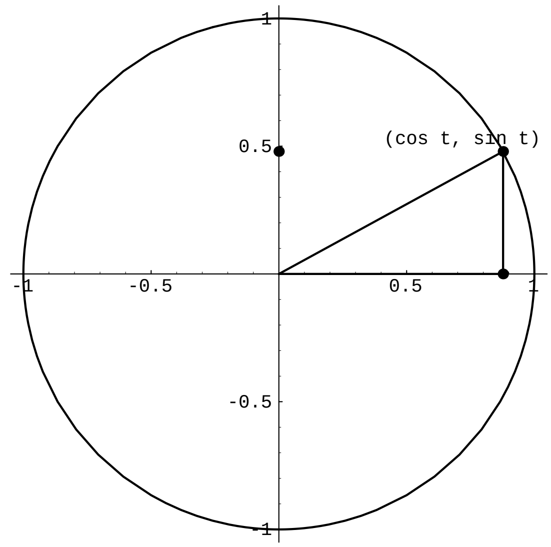

Figure 4.1: The unit circle

Conversely, let t be any real number. We construct the angle with radian measure t. Starting at the point $(1,0)$ we travel the distance $|t|$ along the unit circle (here $|t|$ denotes the absolute value of t). By convention, we travel counter clockwise if t is positive and clockwise if t is negative. In this way we reach a point p on the circle. Let α be the angle between the positive x-axis and the ray which starts at the origin and intersects the unit circle in p. This angle has radian measure t.

Comparison of Angles in Degrees and Radians: We suppose that you are familiar with measuring angles in degrees. The measure of half a revolution (a straight angle) comprises π radians and 180 degrees. So, one degree corresponds to $\pi/180 \approx 0.017453293$ radians, and one radian corresponds to $180/\pi \approx 57.29577951$ degrees. We have the conversion formula

$$x \text{ degrees} = \frac{\pi}{180}\, x \text{ radians}. \tag{4.2}$$

Trigonometric Functions: Let t be once more a real number. Starting at the point $(1,0)$ we travel the distance $|t|$ along the unit circle, counter clockwise if t is positive and clockwise if t is negative. In this way we reach a point $p = (x(t), y(t))$ on the circle, and we set

$$x(t) = \cos t \quad \text{and} \quad y(t) = \sin t. \tag{4.3}$$

This defines the functions $\sin t$ and $\cos t$. You see the construction implemented in Figure 4.1. You can find the graphs of the sine and cosine functions on the interval $[0, 2\pi]$ in Figures 4.2 and 4.3.

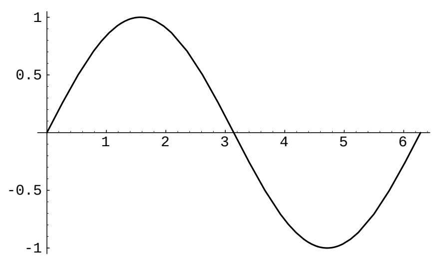

Figure 4.2: $f(x) = \sin x$

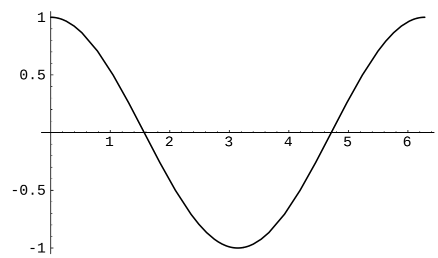

Figure 4.3: $f(x) = \cos x$

The other trigonometric functions, tangent (tan), cotangent (cot), secant (sec), and cosecant (csc) are defined as follows:

$$\tan x = \frac{\sin x}{\cos x} \quad \cot x = \frac{\cos x}{\sin x} \quad \sec x = \frac{1}{\cos x} \quad \csc x = \frac{1}{\sin x} \tag{4.4}$$

To make sure you have some idea about the behavior of the tangent and cotangent function we provided two graphs for each of them. They are drawn over different parts of the domain to show different aspects. See Figure 4.4 to Figure 4.7. You can see the graphs of the secant and cosecant functions in Figure 4.8 and 4.9.

A small table with angles given in degrees and radians, as well as the associated values for the trigonometric functions is given in Table 4.1. If the functions are not defined at some point, then this is indicated by *'n/a'*. Older calculus books may still contain tables with the values of the trigonometric functions, and there are books which were published for the specific purpose of providing these tables. This is really not necessary anymore because any scientific calculator gives those values to you with rather good accuracy.

Trigonometric Functions defined at a right triangle: Occasionally it is more convenient to use a right triangle to define the trigonometric functions. To do this we return to Figure 4.1. You see a right triangle with vertices $(0, 0)$, $(x, 0)$ and (x, y). We may use a circle of any radius r. The

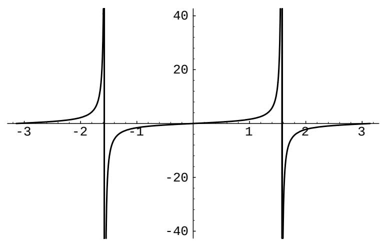

Figure 4.4: $\tan x$ on $[-\pi, \pi]$

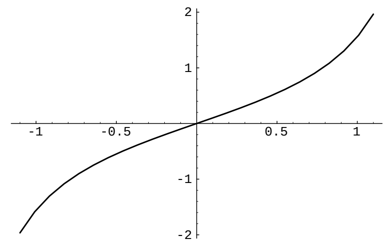

Figure 4.5: $\tan x$ on $[-1.1, 1.1]$

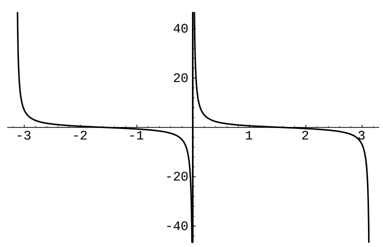

Figure 4.6: $\cot x$ on $[-\pi, \pi]$

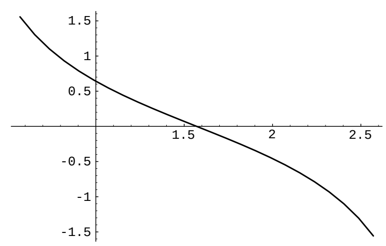

Figure 4.7: $\cot x$ on $[\frac{\pi}{2} - 1, \frac{\pi}{2} + 1]$

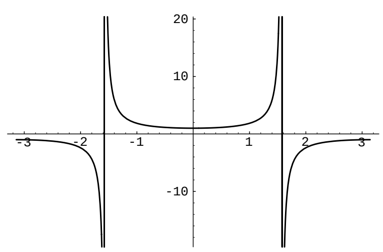

Figure 4.8: $\sec x$ on $[-\pi, \pi]$

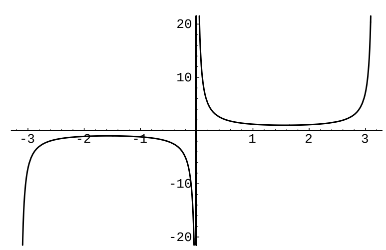

Figure 4.9: $\csc x$ on $[-\pi, \pi]$

degrees	radians	$\sin x$	$\cos x$	$\tan x$	$\cot x$	$\sec x$	$\csc x$
0	0	0	1	0	n/a	1	n/a
30	$\pi/6$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{3}$	$\sqrt{3}$	$\frac{2\sqrt{3}}{3}$	2
45	$\pi/4$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1	1	$\sqrt{2}$	$\sqrt{2}$
60	$\pi/3$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{\sqrt{3}}{3}$	2	$\frac{2\sqrt{3}}{3}$
90	$\pi/2$	1	0	n/a	0	1	n/a
120	$2\pi/3$	$\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$-\sqrt{3}$	$-\frac{\sqrt{3}}{3}$	-2	$\frac{2\sqrt{3}}{3}$
135	$3\pi/4$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	-1	-1	$-\sqrt{2}$	$\sqrt{2}$
150	$5\pi/6$	$\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{3}$	$-\sqrt{3}$	$-\frac{2\sqrt{3}}{3}$	2
180	π	0	-1	0	n/a	-1	n/a

Table 4.1: Values of Trigonometric Functions

right angle is at the vertex $(x, 0)$ and the hypotenuse has length r. Let α be the angle at the vertex $(0, 0)$. In the following the words adjacent and opposing are in relation to α. Then

$$\sin\alpha = \frac{\text{opposing side}}{\text{hypothenuse}} \qquad \cos\alpha = \frac{\text{adjacent side}}{\text{hypothenuse}}$$
$$\tan\alpha = \frac{\text{opposing side}}{\text{adjacent side}} \qquad \cot\alpha = \frac{\text{adjacent side}}{\text{opposing side}}$$
$$\sec\alpha = \frac{\text{hypothenuse}}{\text{adjacent side}} \qquad \csc\alpha = \frac{\text{hypothenuse}}{\text{opposing side}}$$

Trigonometric Identities: There are several important identities for the trigonometric functions. Some of them you should know, others you should be aware of, so that you can look them up whenever needed. From the theorem of Pythagoras and the definitions you obtain

$$\sin^2 x + \cos^2 x = 1, \quad \sec^2 x = 1 + \tan^2 x, \quad \csc^2 x = 1 + \cot^2 x. \tag{4.5}$$

The following identities are obtained from elementary geometric observations using the unit circle.

$$
\begin{array}{rclclcl}
\sin x &=& \sin(x+2\pi) &=& \sin(\pi - x) &=& -\sin(-x) \\
\cos x &=& \cos(x+2\pi) &=& -\cos(\pi - x) &=& \cos(-x) \\
\cos x &=& \sin(x+\frac{\pi}{2}) &=& -\cos(x+\pi) &=& -\sin(x+\frac{3\pi}{2}) \\
\sin x &=& -\cos(x+\frac{\pi}{2}) &=& -\sin(x+\pi) &=& \cos(x+\frac{3\pi}{2})
\end{array}
$$

You should have seen, or even derived, the following addition formulas in precalculus.

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \tag{4.6}$$

$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta \tag{4.7}$$

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \tag{4.8}$$

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \tag{4.9}$$

$$\tan(\alpha+\beta) = \frac{\tan\alpha + \tan\beta}{1 - \tan\alpha\tan\beta} \tag{4.10}$$

$$\tan(\alpha-\beta) = \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta} \tag{4.11}$$

These formulas specialize to the double angle formulas

$$\sin 2\alpha = 2\sin\alpha\cos\alpha \quad \text{and} \quad \cos 2\alpha = \cos^2\alpha - \sin^2\alpha \tag{4.12}$$

From the addition formulas we can also obtain

$$\sin\alpha\sin\beta = \frac{1}{2}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] \tag{4.13}$$

$$\sin\alpha\cos\beta = \frac{1}{2}\left[\sin(\alpha-\beta) + \sin(\alpha+\beta)\right] \tag{4.14}$$

$$\cos\alpha\cos\beta = \frac{1}{2}\left[\cos(\alpha-\beta) + \cos(\alpha+\beta)\right] \tag{4.15}$$

which specialize to the the half-angle formulas

$$\sin^2\alpha = \frac{1}{2}\left[1 - \cos 2\alpha\right] \quad \text{and} \quad \cos^2\alpha = \frac{1}{2}\left[1 + \cos 2\alpha\right] \tag{4.16}$$