Notes & Exercises for Calculus 2009

Raoul Grasman

Chapter 0

Preliminaries

1. Algebra

1.1. Number system. There are four basic sets of numbers that we will use:

$\mathbb{N}$ the Natural numbers: $1, 2, 3, \ldots$,

$\mathbb{Z}$ the Integers: $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$ (from German "Zahlen"),

$\mathbb{Q}$ the Rational numbers: numbers of the form $\frac{n}{m}$, where $n, m \in \mathbb{Z}$, that is, n and m are (*elements of* the) integers, and

$\mathbb{R}$ the Real numbers: the union of the set of rational numbers with the *irrational numbers*. The irrational numbers are numbers for which there are no two integers p and q, such that $\frac{p}{q}$ is exactly that number. Numbers like π and $\sqrt{2}$ are irrational, and hence, $\pi, \sqrt{2} \in \mathbb{R}$, but $\pi, \sqrt{2} \notin \mathbb{Q}$.

Note that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$; that is, these sets are *subsets* ($\subseteq$) of each other. We could add another set of numbers, the so called "complex numbers" $\mathbb{C}$, in which there is a solution to the equation $i^2 = -1$ (that should strike you as impossible! *why?*), but we will not have to deal with them.

Every number can be written in decimal form. E.g., $1/2 = 0.5$ and $1/11 = 0.0909090\overline{09}$, $\pi = 3.14159\ldots$. The line on top of 09 indicates that this pattern is repeated *indefinitely*. The dots after 3.14159 indicates that an infinite amount of decimal numbers follow but that the pattern never starts to repeat. The latter is a hallmark of the irrational numbers. A decimal number is therefore never an exact representation of an irrational numbers, and we therefore prefer to use symbols like π and $\sqrt{2}$ instead.

Example 1. *Give the* exact *solutions to* $x^2 = 2$.

ANSWER. $x = \pm\sqrt{2}$. A calculator gives $x = \pm 1.4142\ldots$ which is not *exact* and therefore *not* the answer to the question. $\square$

Exercise 1. *Give the exact solution to* $x^2 = 3$.

Given the fact that an irrational number can be approximated by a decimal number to any desired accuracy (simply by taking enough decimal places into account), it is not surprising that any real number (in $\mathbb{R}$) can be approximated arbitrarily closely by a rational number (in $\mathbb{Q}$).

It comes to many as a surprise, and strikes most people as counter intuitive that the number 1 is equivalently written $0.999\overline{9}$—that is, they are one and the same number.

Exercise 2. *Suppose they were different. Then it should be possible to find a positive number such that* $0.999\overline{9}$ *plus that number equals* 1. *Can you come up with a number that you can add to* $0.999\overline{9}$ *such that the result is* 1? Hint: *Start with* 0.001 *and try ever smaller values.*

1.2. Algebra. In algebra, arithmetic with numbers is replaced by arithmetic with letters. The basic algebra rules you learned in high school are
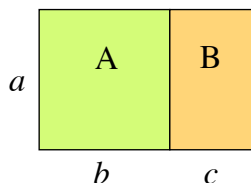

FIGURE 1. Land measurement (geo-metry) motivation for algebra rules.

$$
\begin{aligned}
a + b &= b + a \\
(a + b) + c &= a + (b + c) \\
ab &= ba \\
(ab)c &= a(bc) \\
a(b + c) &= ab + ac
\end{aligned}
$$

Some of these rules are motivated in the rectangle in figure 1; they are taken as *axioms*—i.e., *assumed* true statements—in algebra.

Subtraction & division: inverse operations. Subtraction and division are define in terms of addition and multiplication: For the equation

$$x + 3 = 10$$

we may ask "for what number x is this statement (equation) true?" This is the opposite of asking "how much is 7 plus 3?" This is an example of an "inversion". Similarly, asking "for what number x is

$$3x = 6$$

true?" is an inversion of the question "how much is 3 times 2?". The inverse operations for addition is denoted by $x = 10 - 3$ of course, and for multiplication by $x = 6/3$.

Inversions may be though of as questions of how to *undo an operation.* Inversions, or the idea of an inverse of an operation (e.g., the inverse of multiplying by a number or of adding a number) are in many ways central to mathematical analysis.

The inverse of an operation in not always defined. For instance, there is no x satisfying $0 \cdot x = 10$, hence the notation $10/0$ is meaningless—you may not divide by zero. As another example, consider the multiplication of c by a again as the area of the rectangle B in figure 1. Suppose we let a grow larger and larger, while c becomes smaller and smaller, but we make sure that always $a \cdot c = 10$. In the limit, we symbolically get the form

$$\lim_{a \to \infty, c \to 0} a \cdot c = \text{“}\infty \cdot 0\text{”} = 10,$$

but obviously, the number 10 was completely arbitrary and so the symbol "$\infty \cdot 0$" has no meaning. More subtly, $0/0$ is not defined, nor is ∞/∞.

Yet for finite $a \neq 0$ it is useful to think of $a/0$ as ∞ or $-\infty$, depending on the sign of a, since a/x approaches these limits as $x \to 0$, i.e.,

$$\lim_{x \to 0} \frac{a}{x} = \pm\infty.$$

(Note we must assume in this rule of thumb that x approaches 0 from above, because otherwise the sign would be flipped.) Saying however that $\infty/\infty = 1$ is an error. Similarly $\infty - \infty$ is not defined, and saying $\infty - \infty = 0$ is an error.

Quadratic relations. Some basic results from high school algebra are

$$
\begin{aligned}
(a+b)^2 &= a^2 + 2ab + b^2, \\
(a-b)^2 &= a^2 - 2ab + b^2, \\
(a+b)(a-b) &= a^2 - b^2,
\end{aligned}
$$

which are helpful in solving polynomial equations of the form

$$ax^2 + bx + c = 0.$$

Solving this equation means "finding those values of x for which this statement is true (the equation holds)". If the *discriminant* $D = b^2 - 4ac \geq 0$ the general solution to this equation is

$$x_{\pm} = \frac{-b \pm \sqrt{b^2 - 4\,a\,c}}{2a}.$$

Otherwise, if the discriminant $D < 0$, the equation has no solutions for x in $\mathbb{R}$ (but in $\mathbb{C}$, which we won't discuss).

Exponents. For repeated multiplication by the same number a we have the shorthand notation

$$\overbrace{a \cdot a \cdots \cdots a}^{n\times} = a^n.$$

Here the *exponent* n is a natural number (i.e., $n \in \mathbb{N}$), because in terms of the shorthand anything else doesn't make sense (yet! *why?*). We can however define

$$a^{-n} = \frac{1}{a^n},$$

which is consistent with the above shorthand (why?). The rules for calculating with exponents are

$$a^n \cdot a^m = a^{n+m}, \quad (ab)^n = a^n b^n, \quad a^{-n} = \frac{1}{a^n} = \left(\frac{1}{a}\right)^n, \quad (a^n)^m = a^{nm} = (a^m)^n.$$

We can also ask "for which number x is $x^3 = 125$ true?". This is again an example of an inversion. It is mostly a lot of work to compute the solution and so we use a machine for this. The answer is of course the cubic root of 125, written $\sqrt[3]{125}$, but for general cases we may define the notation $125^{1/3}$ to mean just that. If we do, then in general if $b = a^m$ then if we ('formally') apply the last rule for exponents $b^{1/m} = (a^m)^{1/m} = a^{m \cdot 1/m} = a^1 = a$ is exactly what we were looking for: "the value of x for which $x^m = b$", namely a. Raising a number a to the power of $1/n$ is the *inverse operation* of raising a to the power of n (which was initially only a shorthand).

Note for integer m and n, using the rules so far obtained, we have $(a^n)^{1/m} = a^{n/m}$. Because $n/m \in \mathbb{Q}$, i.e., n/m is a rational number, the shorthand notation is now defined for any exponent that is a rational number! Well, it is true as long as $(a^n)^{1/m}$ and $(a^{1/m})^n$ make sense, that is, if for instance m is even a cannot be negative (for example $(-4)^{1/2} = \sqrt{-4}$ is not a real number, while $(-8)^{1/3} = \sqrt[3]{-8} = -2$ is). If you remember from two pages ago that any real number can be arbitrarily closely approximated by a rational number, you can imagine that it may be possible to

extend the definition of a^x from only rational numbers x to all real numbers x. This is the case, but only if $a \geq 0$, Then the same rules are valid:

$$a^x a^y = a^{x+y}, \quad (ab)^x = a^x b^x, \quad , a^{-x} = \left(\frac{1}{a}\right)^x, \quad (a^x)^y = a^{xy} = (a^y)^x, \quad x, y \in \mathbb{R}.$$

Two additional rules are special cases:

$$0^x = 0, \quad \text{and} \quad x^0 = 1, \text{ for any } x \in \mathbb{R}, \text{ in particular } 0^0 = 1.$$

Another thing we can ask, another inversion of the shorthand notation, is the question of "for what value of x is $a^x = b$ true?".

Exercise 3. *What's the difference between this question and the previous?*

The answer to the question posed is of course the base a logarithm of b, $\log_a b$. Like A numerical answer is also not easy to compute for a human, and although a machine can do so, it is most of the time completely unnecessary and not even useful to actually compute their value. We can simply denote the answer symbolically as

$$\log_a b,$$

and be satisfied. We do need some rules for computing with logarithms.

The most important rules are

$$\log_a(xy) = \log_a x + \log_a y, \quad \log_a(x^y) = y \log_a(x).$$

The first is from the fact that by definition $a^{\log_a(xy)} = xy = (a^{\log_a x})(a^{\log_a y}) = a^{\log_a x + \log_a y}$, which is a consequence of the first rule for exponents. The second rule follows from the first if y is a natural number:

$$\log_a x^y = \log_a(\overbrace{x \cdot x \cdots x}^{y\times}) = \overbrace{\log_a x + \log_a x + \cdots + \log_a x}^{y\times} = y \log_a x.$$

It is also true for more general y however; we will return to this later.

The algebra rules recalled so far, and the inversion of the various operations, are essential to all of mathematics. Any successful attempt to learn calculus requires you to become very well skilled in applying them. The next set of exercises are intended to drill you.

1.3. Exercises. This section contains some exercises on algebra.

Example 2. *Solve the following equation for x:*
$$\frac{a^m x^2}{\log_a(b)} = b^3 \pi.$$
Assume that a is not an integer.

ANSWER. We need to rewrite the equation so as to undo all the operations that have been done on x (we need to solve for x). We first do by checking at each step if the inverse operation is allowed, and under what condition.

Lets first do the inverse operation of division by $\log_a(b)$ (namely multiplication by that number on both sides of the equation). We may safely assume that $\log_a(b)$ is not 0, because otherwise the equation doesn't make any sense. Hence, the first step in rewriting is

$$a^m x^2 = b^3 \pi \log_a(b).$$

Next, lets divide through by a^m. to be able to do so, a^m should not be zero, that is $a \neq 0^{(}1/m)$ should not be zero. Then

$$x^2 = \frac{b^3 \pi \log_a(b)}{a^m}.$$

Next, we consider taking roots to undo the square. First of all the equation can only be true if the right hand side is positive or zero. Because of the logarithm, a and b must be positive, unless a is an integer, otherwise the logarithm isn't defined. The question states that a is not an integer, so the right hand side is positive only if $\log_a(b) \geq 0$. To find a condition on b in this case we raise a to the power of both sides of the equation:

$$a^{\log_a(b)} \geq a^0 \iff b \geq 1,$$

hence $b \geq 1$ should be the case. Assuming this, we can safely take the square root of the right hand side, but we should be careful to note that there are two solutions:

$$x = \pm\sqrt{\frac{b^3 \pi \log_a(b)}{a^m}}.$$

$\square$

Exercise 4. *Solve the equation for x. That is, give x if it is assumed that the equation is true. Indicate possible constraints on the unspecified variables. You may assume that the base of a logarithm is never an integer. Simplify the expression as much as possible.*

(1) $ax + b = 0$.

(2) $ax^2 + b = 0$.

(3) $a^m x = b$.

(4) $a^m x = -b^n \pi$

(5) $\log_b(x) = a$

(6) $ax^{2m} = b^{2n} \dfrac{\sqrt{2}}{\pi^{m/2}}$

(7) $\dfrac{a^b x^m}{\sqrt{\pi}} = \dfrac{y^b}{m}$

(8) $\log_a(x^b y) = \pi$

(9) $\log_b(x^a y^{\log_b y}) = \sqrt{\pi^2}(\log_b y)^2$

(10) $x^{\log_a(a^{m-n})} \log_a(y^m) = \dfrac{\pi}{mx^n y^m}$

(11) $\log_b(4x^2 + 4x + 1) = y^2$

2. Functions

Functions are mappings between a points in one set, the *domain*, and points in another, the *range*. Each point in the domain should be associated with one *and only one* point in the range, otherwise it is not a function. Points in the range may be associated with multiple points however.
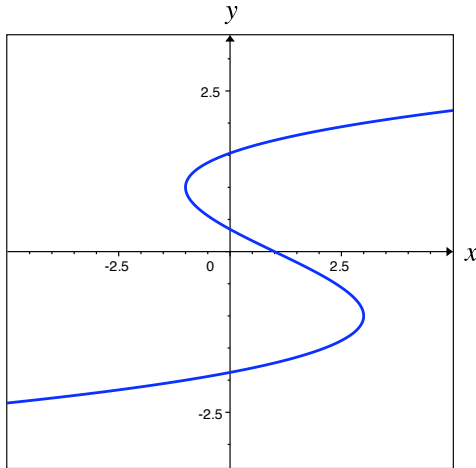
The following are functions:
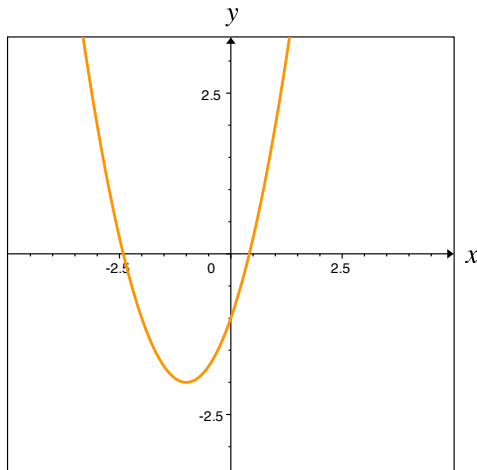
x	$g(x)$
-3	65
-2	5
-1	-5
0	-1
1	5
2	25

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

The graph is a function if the x-axis is the input set and the y-axis the output set. The function $g(x)$ is defined only on a discrete set of values. The function $f(x)$ implicitly has all real numbers as its input set.

The following are not functions (*why?*):

$f(x) = y$ where $y^2 + x^2 = 3$.

Polynomial functions. One of the most important class of functions are the *polynomials*. These are functions of the form

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

where n is an integer, and $a_n, a_{n-1}, \ldots, a_1, a_0$ are real valued fixed coefficients. For example

$$f(x) = 4x^2 + 4x + 1$$

is a quadratic polynomial, while

$$g(x) = x^3 - 3$$

is a cubic polynomial function. In general they are called n-th order polynomial if the highest power in x is n. The polynomial is of course completely determine by its coefficients, but alternatively, it is also completely determined by its so called *roots* (or *zero's*). The *roots* of a polynomial are the set of values of x for which $f(x)$ satisfies the equation

$$f(x) = 0.$$

(Strictly, this is only true if we allow x to assume also complex values, i.e., $x \in \mathbb{C}$; if we insist on $x \in \mathbb{R}$, we can be sure that $f(x)$ has no more than n roots.) If $x_1, x_2, x_3, \ldots$ are the roots of a

polynomial $f(x)$, then $f(x)$ can be written

$$f(x) = (x - x_1)(x - x_2)(x - x_3) \cdots .$$

If we allow complex roots, an n-order polynomial has exactly n roots. This fact is known as the *Fundamental Theorem of Algebra*. Knowing the roots of a polynomial is far more useful than knowing its coefficients. Coefficients are easy to compute from the roots, but not vice versa. Polynomials show up in many contexts. They will reappear in approximation, numerical integration, Taylor expansions and Taylor series.

One-to-one functions. Some function are one-to-one:[1] For each value y in the range of f, there is one and only one x such that $y = f(x)$. Some examples help.

Example 3. *The function $f(x) = x^2$ is not one-to-one because for $f(x) = 4$ I can find two values of x for which this is true, namely $x = 2$ and $x = -2$. In this case there are two originals for that same image.*

Example 4. *The function $f(x) = ax + b$ is one-to-one if $a \neq 0$, because the equation $f(x) = 4$ is only true for $f(x) = 4 \iff 4 = ax + b \iff 4 - b = ax \iff x = (4 - b)/a$ because $a \neq 0$. The number 4 here was arbitrary, the same is true for any other value.*

Example 5. *The function $f(p) = p/(1 - p)$, $p \neq 1$ is one-to-one, because for y in the range of $f(p)$, $f(p) = y$ is only true if $y = p/(1 - p) \iff y(1 - p) = p \iff y - py = p \iff y = p + py \iff y = p(1 + y) \iff p = y/(1 + y)$.*

Example 6. *The function in the graphs below is one-to-one, because if we exchange the x and y-axis by flipping in the diagonal (right graph), we obtain the graph of a function (that is, for each input there is only one output)*

---

[1]strictly, here we mean one-to-one *and* onto
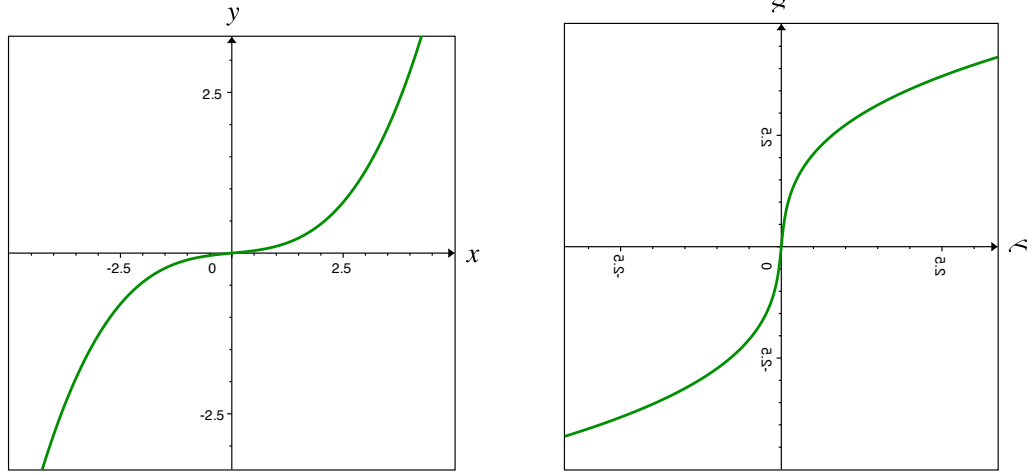
As you may have noticed from these examples, to check if a function is one-to-one we need to take an arbitrary value y that $f(x)$ can assume and solve the equation

$$y = f(x)$$

for x. That is, we need to undo (invert) all the operations that f does to x. Stated differently, we need to find the function that is the inverse operation of f. If the function g constitutes the inverse operation of f, that is, if g undoes all operations carried out by f on x, then clearly

$$g(f(x)) = x,$$

because for any value y that $f(x)$ may assume, $g(y)$ returns the value of x for which $y = f(x)$ is true, hence $x = g(y) = g(f(x))$. The reverse is also true: $f(g(y)) = y$, because in $f(x) = y$ we can replace x by $g(y)$, hence $y = f(x) = f(g(y))$. This is hallmark of inverse functions: if g is the inverse function of f then also is f the inverse function of g.

Because inverse functions come in pairs, they are often denoted f and f^{-1}, which is reminiscent of the shorthand a^n notation for multiplying n times by a number a. In that case, the inverse function is multiplying by a^{-n}, since if $y = a^n x$, then $a^{-n} y = a^{-n}(a^n x) = a^{-n+n} x = x$, and so we have solved for x. Analogously, $f^{-1}(f(x)) = x$. This example seems to be responsible for the often encountered mistake that the inverse of an arbitrary function $f(x)$ is thought to be

$$\text{``}f^{-1}(x) = \frac{1}{f(x)}\text{''}, \quad \text{\textcolor{red}{this is wrong!}}$$

which is clearly incorrect if we consider the concept behind the notation. Adding to the confusion is the notation $f(x)^{-1}$, which *does* mean

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

The notation $f^{-1}(f(x)) = x$ and $f(g(y)) = y$ leads us to the consideration of *composit functions*. Composite functions are in essence functions of the outcome of other functions. This makes it sound special, and it is totally not. Any function is a composite function. Any function carries out one or more operations on an input value x. For example the function

$$f(x) = \log_a(3x^2 + x)$$

multiplies x by itself (x^2), multiplies the result by 3 ($3x^2$), adds x to the outcome ($3x^2 + x$), and takes the base a logarithm of that value ($\log_a(3x^2 + x)$). That is, it does four "elementary" operations.[2]

Each of these operations can constitute a function by itself. For instance, using the above decomposition, we can write f as

$$f(x) = g(h(x)),$$

where

$$g(y) = \log_a(y),$$

and

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

But we could also decompose f into

$$f(x) = g(h(x)), \qquad \text{where } g(y) = \log_a\left(3y^2 - \frac{1}{12}\right), \text{ and } h(x) = x + \frac{1}{6}.$$

You should not only be able to verify this, but also be able to find decompositions yourself. One should be careful however: The decomposition $g(y) = \log_a(3y + \sqrt{y})$, $h(x) = x^2$ may seem valid, but verification shows that $g(h(x)) = \log_a(3h(x) + \sqrt{h(x)}) = \log_a(3x^2 + \sqrt{x^2}) = \log_a(3x^2 + |x|) \neq \log_a(3x^2 + x) = f(x)$.

2.1. Exercises. The exercises in this section concern inverse functions and composite functions.

Example 7. *Find the inverse of*

$$f(x) = \left(\frac{2x}{3x^2 + x}\right)^m.$$

ANSWER: We need to solve the equation

$$y = f(x) = \left(\frac{2x}{3x^2 + x}\right)^m.$$

First, take the m-th root on both sides,

$$y^{1/m} = \frac{2x}{3x^2 + x}.$$

Notice the common factor x in the numerator and denominator on the right hand side, it cancels,

$$y^{1/m} = \frac{2x}{x(3x + 1)} = \frac{2}{3x + 1}.$$

Now invert enumerator and denominator on both sides,

$$\frac{1}{y^{1/m}} = \frac{3x + 1}{2},$$

and subtract 1/2 from both sides,

$$\frac{1}{y^{1/m}} - \frac{1}{2} = \frac{3x}{2};$$

multiply by 2/3,

$$\frac{2}{3}\frac{1}{y^{1/m}} - \frac{1}{3} = x.$$

Therefore, $f^{-1}(y) = \frac{2}{3}y^{-1/m} - \frac{1}{3}$.   □

Exercise 5. *Find the inverses of the following functions.*

---

[2]Of course it is difficult to define which operations are "elementary".

(1) $f(x) = ax + b$

(2) $f(x) = \frac{ax}{b-x}$

(3) $f(x) = \log_a(x^b - c) + \pi$

(4) $f(x) = \dfrac{ab^x}{c - b^x}$

(5) $f(x) = \dfrac{ab^x}{c - b^{x-c}}$

(6) $f(x) = \dfrac{x-a}{x-b}$

(7) $f(x) = \sqrt{\log_a(b^x + c)}$

(8) $f(x) = \left(\dfrac{2x-b}{3x-a}\right)^m$

Example 8. *Find functions $g(y)$ and $h(x)$ such the $f(x)$ is the composite function $g(h(x))$, where*

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

ANSWER: An obvious decomposition is $g(y) = \frac{y+1}{-y}$ and $h(x) = x^3 - 5x^2$, then

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

A less obvious decomposition is

$$f(x) = \frac{x^3 - 5x^2 + 1}{-x^3 + 5x^2} = \frac{x^3 - 5x^2 + 1}{1 - x^3 + 5x^2 - 1} = \frac{(x^3 - 5x^2 + 1)}{1 - (x^3 - 5x^2 + 1)} = g(h(x)),$$

where

$$g(y) = \frac{y}{1-y}, \quad \text{and } h(x) = x^3 - 5x^2 + 1.$$

□

Exercise 6. *Find at least two ways to write $f(x)$ as a composite function.*

(1) $f(x) = \dfrac{a}{x-b}$

(2) $f(x) = \dfrac{-x}{x+a}$

(3) $f(x) = \dfrac{x-a}{x^2+10}$

(4) $f(x) = \log_a\left[\left(\dfrac{x}{1-x}\right)^m\right]$

(5) $f(x) = \dfrac{1}{\sqrt{1/x^2 - 1}}$

(6) $f(x) = \left(\sqrt{\log_a(x^2 + b/a)}\right)^n$

Applied Exercises 1. *The following exercises concern applications of functions.*

(1) *A psychophysical experiment was conducted to analyze human response to electrical shocks. The subjects received a shock of a certain intensity. They were told to assign a magnitude of 10 to this particular shock, called the standard stimulus. Then other shocks (stimuli) of various intensities were given. For each one the response R was to be a number that indicated the perceived magnitude of the shock relative to that of the standard stimulus. It was found that R was a function of the intensity I of the shock (I in microamperes) and was estimated by $R = f(I) = I^{4/3}/2500$, $500 \le I \le 3500$.*
   (a) *Evaluate $f(1000)$.*
   (b) *Evaluate $f(2000)$.*
   (c) *Suppose that I_0 and $2I_0$ are in the domain of f. Express $f(2I_0)$ in terms of $f(I_0)$. What effect does the doubling of intensity have on the response?*

(2) *In a paired-association learning experiment, the probability of a correct response as a function of the number n of trials has the form $P(n) = 1 - 0.5(1-c)^{n-1}$, $n \ge 1$, where the estimated value of c is 0.344. Find $P(1)$ and $P(2)$ by using this value of c.*

(3) *Studies have been conducted concerning the statistical relations between a person's status, education and income. Let S denote a numerical value of status based on annual income I. For a certain population, suppose that $S = f(I) = 0.45(I - 1000)^{0.53}$. Furthermore, suppose that a person's income I is a function of the number of years of education E, where $I = g(E) = 7202 + 0.29 \times E^{3.68}$. Find $(f \circ g)(E)$. What does this function describe?*

(4) *In an experiment on visual information processing, a subject briefly viewed an array of letters and was then asked to recall as many letters from the array as possible. The procedure was repeated several times. Suppose that y is the average number of letters recalled from arrays with x letters. The graph of the results approximately fits the graph of $y = f(x) = x$, if $0 \le x \le 4$, $f(x) = x/2 + 2$, if $4 < x \le 5$, $f(x) = 4.5$, if $5 < x \le 12$. Graph this function.*

(5) *For reasons of comparison, a psychometrician wants to rescale the scores on a set of test papers so that the maximum score is still 100 but the mean (average) is 80 instead of 56. Find a linear equation that will do this. [Hint: You want 56 to become an 80 and 100 to remain 100. Consider the points (56, 80) and (100, 100) and, more generally, (x, y) where x is the old score and y is the new score. Find the slope and use a point-slope form. Express y in terms of x. If 60 on the new scale is the lowest passing score, what was the lowest passing score on the original scale?*

(6) *The result of Sternberg's experiment on information retrieval is that a person's reaction time R, in milliseconds, is statistically a linear function of memory set size N as follows: $R = 38N + 397$. Sketch the graph for $1 \le N \le 5$. What is the slope?*

(7) *In a certain learning experiment involving repetition and memory, the proportion p of items recalled was estimated to be a linear function of effective study time t (in seconds), where t is between 5 and 9 inclusive. For an effective study time of 5 seconds, the proportion of items recalled was 0.32. For each 1-second increase in study time, the proportion recalled increased by 0.059.*
   (a) *Find an equation that gives p in terms of t.*
   (b) *What proportion of items was recalled with 9 seconds of effective study time?*

3. Miscellaneous

Binomial coefficient.

Systems of equations.

Linear.
Nonlinear.

Chapter 1

Derivatives

This chapter introduces the derivative of a function and the technique(s) of differentiation of functions. Because the existence of the derivative of a function requires that the function is continuous, we first need to know what we mean by continuity and what a continuous function is. To make the idea of continuity precise, we need to know a bit about limits. But only a bit.

1. Continuity of functions and limits

Some functions have sudden jumps, while other functions display only "smooth" changes as the input changes smoothly. The former type of functions are (called) *discontinuous functions*, while the latter are *continuous*. Other types of discontinuous functions are functions not defined at certain points, or with asymptotes. Intuitively, a continuous function is a function whose graph can be drawn "without lifting the pen from the paper".

Example 9. *Here are examples of a continuous function, and of three types of discontinuous functions:*

$$g(x) = ax^3 + bx - b, \qquad f(x) = \begin{cases} ax^2 + b & x > d \\ ax - b & x \le d \end{cases}, \qquad u(x) = \frac{1}{x-2}, \qquad v(x) = \frac{x^2 - x - 2}{x - 2}.$$

They are graphed below. The solid end point at $x = -2$ on $f(x)$ indicates the $x \le -2$ part (-2 is inclusive for the segment) of the definition of f, while the open circle at $x = -2$ indicates the $x > -2$ part (-2 is exclusive for the segment). The function $u(x)$ has a discontinuity at $x = 2$ because of the asymptote at $x = 2$. The function $v(x)$, while in fact a straight line, is undefined at $x = 2$ where it is discontinuous.
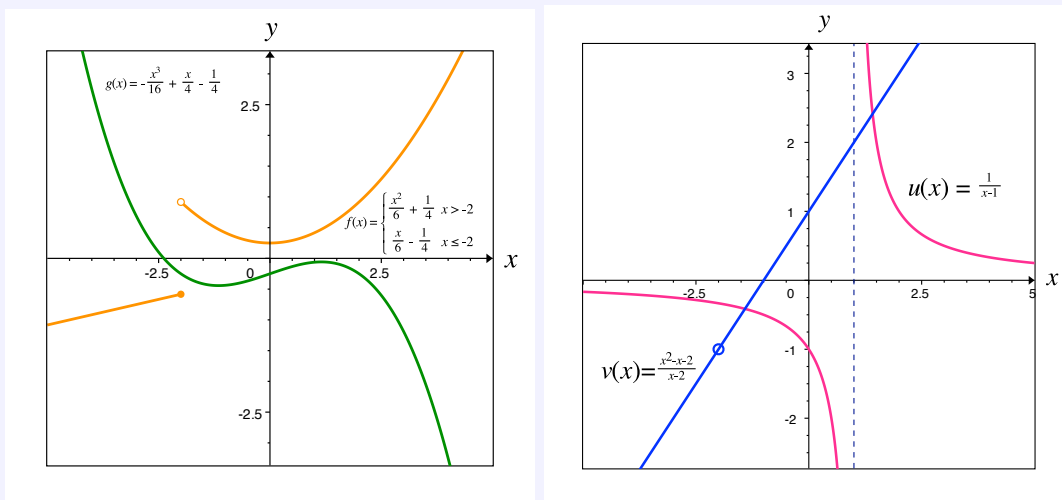
To give a (more) precise definition of continuity, we need *limits*.

The limit of a function $f(x)$ at $x = a$ is that unique number L that $f(x)$ will reach arbitrarily closely if we let x approach a from the left or right arbitrarily closely without x ever becoming exactly a. The mathematical notation is

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

Of course, L depends on the value of a.

For example, the limit of $h(x) = 2x + 1$ as $x \to 1$ is 3, as $h(x)$ gets arbitrarily close to 3 as x gets arbitrarily close to 1. On the other hand, 3.001 is *not* the limit because $h(x)$ does not get closer to 3.001 as x gets closer to 1 (in fact, if we start at $x = 3.0005$, $h(x)$ gets farther away from 3.001 as $x \to 1$).

This may seem an awkward, circuitous definition that is difficult to understand. But the importance lies in the fact that some functions are not defined at a certain point, or are such that the limit at a certain point does not exist.

Example 10. *Consider the function*

$$h(x) = \frac{\sin(x)}{x}.$$

This function is not defined at $x = 0$ (why?). Yet, take a look at the graph of this function depicted below. Clearly, however close we take x to 0, the $h(x)$ gets closer and closer to 1. Hence,

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1$$

Example 11. *Consider the function $f(x)$ of Example 9. We look at the "interesting" point $x = -2$. Note that if x approaches -2 from the* right*, then the limit seems to be 22/24, because $(-2)^2/6 + 1/4 = 22/24$, but if x approaches -2 from the* left*, the limit seems to be $(-2)/6 - 1/4 =$*
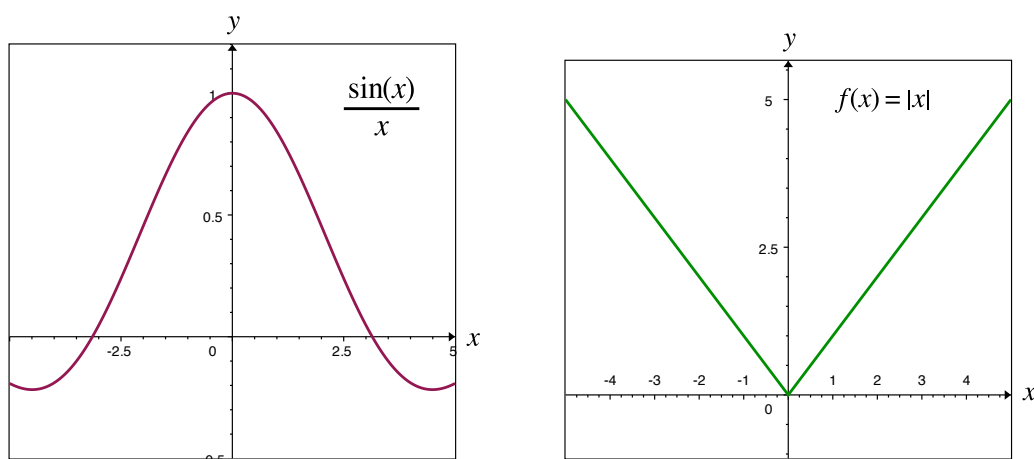
$-7/12$. *And so there are two numbers that f approaches when $x \to -2$, depending on the direction from which x approaches -2. Because the limit should be unique, the limit of $f(x)$ at $x = -2$ does not exist! (At other points the limit does exist however. E.g., $\lim_{x \to 0} f(x) = 1/4$.)*

Example 12. *Consider the function*

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

depicted above. We look at the interesting point $x = 0$. The function is continuous in this point. The function is continuous in this point, because if x approaches 0 either from the right or from the left it gets arbitrarily close to 0, and $f(0) = 0$ by definition of $|x|$.

The limit of a function can be finite, or infinite ($\pm\infty$), but not necessarily exists:

Example 13. *The limit of $f(x) = \sin(1/x)$ as $x \to 0$ does not exist, because $\sin(y)$ keeps oscillating between -1 and 1, no matter how large $y = 1/x$ becomes.*

Calculating limits can be quite difficult. For instance, the limit of $\sin(x)/x$ in Example 10 is not so easy to establish. Merely looking at the graph as we did is not a proof. For some functions it suffices to evaluate the function in the point a and see if it is defined (that is, it doesn't result in expressions such as $0 \cdot \infty$, $\infty - \infty$, ∞/∞). This should always be tried. Although we will not pursue calculating limits here, we list some rules that hold for limits.

If $f(x)$ and $g(x)$ are functions for which the limits $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist, then

(1) $$\lim_{x \to a} c \cdot f(x) = c \cdot \lim_{x \to a} f(x) \qquad \text{(constant multiple)}$$

(2) $$\lim_{x \to a} [f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x) \qquad \text{(sum)}$$

(3) $$\lim_{x \to a} [f(x)\,g(x)] = \lim_{x \to a} f(x) \lim_{x \to a} g(x) \qquad \text{(product)}$$

If, in addition, $\lim_{x \to a} g(x) = b$ and $f(x)$ is continuous in b, then $\lim_{x \to a} f(g(x)) = f(b)$.

Continuity of a function can now be given a precise definition with help of limits: A function $f(x)$ is continuous in a point a if the limit of $f(x)$ as $x \to a$ is equal to $f(a)$, i.e.,

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

Example 14. *The function $h(x) = \sin(x)/x$ of Example 10 is not continuous in the point $x = 0$, because*

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1 \neq \frac{\sin(0)}{0} = \frac{0}{0},$$

which is not defined.

A function is called *uniformly continuous* if it is continuous for all x in the domain of the function.

2. Derivatives

Secant approximation. To approximate the behavior of a function $f(x)$ we can use a secant approximation: Take two points on the horizontal axis, a and b, with $a < b$, and take as the secant approximation the straight line that cuts the function $f(x)$ in a and b. The left panel of figure 1 illustrates this. The straight line is given by the equation

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

Using your geometric intuition, you should be able to derive this equation for yourself. Using this secant approximation we can fairly accurately predict how the value of $f(x)$ increases of the course of x running from a to b.
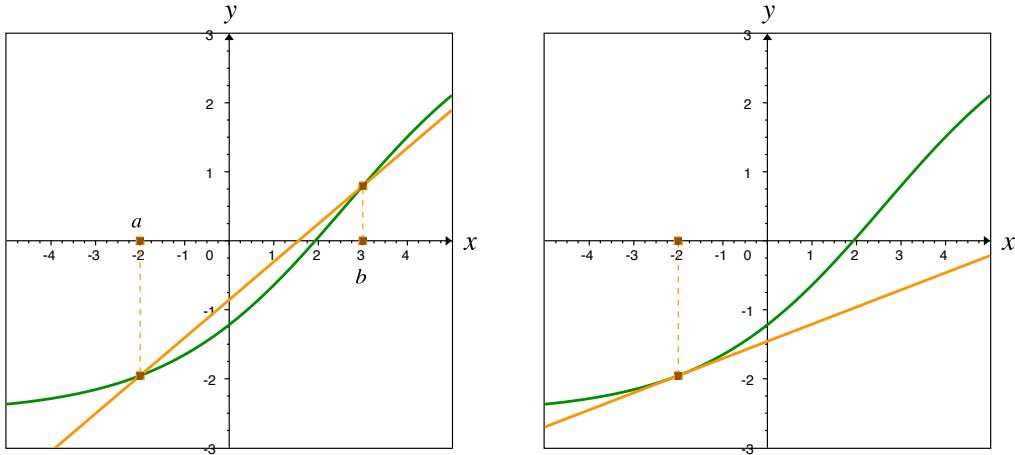

FIGURE 1. Left: Secant approximation. Right: Tangent approximation. The green line is $f(x)$ the straight orange line is the secant approximation using the points $a = -2$ and $b = 3$.

The quality of the approximation of course depends on how strongly the function bends on the line segment (a, b). However, whenever a function is continuous, if we take the interval (a, b) small enough (we "zoom in"), the curve of the function looks more an more like a straight line, and in the limit that $b \to a$, the secant line becomes an excellent approximation to $f(x)$ in the interval (a, b) and is then called the *tangent line* of the curve at a. This is illustrated in the right panel of figure 1. The coefficient,

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

the *slope* of the secant line, measures the steepness of the line, and hence, indicates how fast $f(x)$ increases or decreases over the interval (a, b)—it indicates the *rate of change* of $f(x)$.

Derivative. The *derivative* of $f(x)$ is defined to be this limit of this secant line coefficient. That is, the derivative of $f(x)$ at a point a is defined as

$$f'(a) = \lim_{b \to a} \frac{f(b) - f(a)}{b - a}, \qquad \text{or, equivalently,} \qquad f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}.$$

(*Why are the two equivalent?*) It indicates the change of $f(x)$ due to a small increase in x relative to size of the increase in x at a specific point a on the horizontal axis. Note that this limit does not necessarily exist for every possible point a, and so the derivative may not exist at certain points. If the limit exists in a point a, $f(x)$ is called *differentiable* in that point. If the derivative of $f(x)$ exists in any point a, we say that $f(x)$ is *differentiable*. It turns out that the limit can only exist if $f(x)$ is continuous in a. A differentiable function is always continuous (but not necessarily vice versa). When we talk about "the derivative of $f(x)$" we usually assume that it is a continuous function of a. If $f'(a)$ is a continuous function of a, then $f(x)$ is called *continuously differentiable*.

Example 15. *We calculate the derivative of the function $f(x) = x^2$.*

$$\begin{aligned}
f'(x) &= \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} & \text{(by definition)} \\
&= \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h} \\
&= \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\
&= \lim_{h \to 0} \frac{2xh + h^2}{h} \\
&= \lim_{h \to 0} 2x + h \\
&= 2x.
\end{aligned}$$

Hence, $f'(x) = 2x$.

The tangent line approximation to $f(x)$ is analogous to the secant line, except that the derivative of f is used. The equation for the tangent at a point a is

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

Exercise 7. *Verify that $f'(x) = 2x$ gives the slope of the tangents to $f(x) = x^2$ at any point x on the graph. Use your favorite computer program to draw the function $f(x)$ and draw the tangents at a number of different locations x.*

We have used a prime (′) to indicate the derivative of $f(x)$. Other notations are

$$\frac{d}{dx} f(x), \frac{df}{dx}, \text{ and } y'.$$

In physics you sometimes see $\dot{y}$ to indicate time derivatives.

If the derivative $f'(x)$ of $f(x)$ is a function (as we usually assume), it *too* may be differentiable. The derivative of $f'(x)$ is called the *second order* derivative of $f(x)$, often denoted

$$f''(x), \frac{d^2}{dx^2} f(x), \frac{d^2 f}{dx^2}, y'', \text{ or } f^{(2)}(x).$$

The third order derivative of $f(x)$ is denoted $f'''(x)$, $\frac{d^3}{dx^3}f(x)$ etc. Higher order derivatives are indicated only with $\frac{d^n}{dx^n}f(x)$ and $f^{(n)}(x)$, where n is the order. Parentheses are used to distinguish with the notation $f^n(x) = f(f(f(\cdots f(x))))$ and $f(x)^n = f(x)\cdots f(x)$. (Note that $n \in \mathbb{N}$—given the current definition this is the only thing that makes sense.)

Because finding limits is not always easy (or requires a lot of work), finding derivatives from the definition in terms of a limit is not always easy (or requires a lot of work). Fortunately, Calculus provides the rules to calculate derivatives (relatively) easily, by braking it down in terms of derivatives of simpler, easier to differentiate functions.

We simply list these rules

(4)	$[c \cdot f(x)]' = c \cdot f'(x)$	(constant multiple rule)
(5)	$[f(x) + g(x)]' = f'(x) + g'(x)$	(sum rule)
(6)	$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$	(product rule)
(7)	$\left[\dfrac{f(x)}{g(x)}\right]' = \dfrac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$	(quotient rule)
(8)	$[f(g(x))]' = f'(g(x))g'(x)$	(chain rule)

The last rule is the most confusing, but also the most powerful. It is often written in the suggestive form

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}.$$

One should keep in mind with this notation however that the function $\frac{df}{dg}$ has to be evaluated in $g(x)$, and the function $\frac{dg}{dx}$ in x.

Example 16. *To verify the first rule, we use the limit definition of derivatives and the limit rules: Suppose $g(x) = c \cdot f(x)$ and suppose we know $f'(x)$. Then:*

$$\lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \to 0} \frac{c \cdot f(x+h) - c \cdot f(x)}{h} \qquad \textit{(by definition of derivative)}$$

$$= c \cdot \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \qquad \textit{(by limit rule 1)}$$

$$= c \cdot f'(x) \qquad \textit{(by definition of derivative)}$$

Example 17. *Verifying the third rule differentiation rule requires the zero trick. Note that if $u(x) = f(x)g(x)$, then $u(x+h) = f(x+h)g(x+h)$.*

$$u'(x) = \lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}$$

$$= \lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h} \qquad \textit{(zero trick)}$$

$$= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}g(x+h) + \lim_{h \to 0} f(x)\frac{g(x+h) - g(x)}{h} \qquad \textit{(limit rule 2)}$$

$$= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}\lim_{h \to 0} g(x+h) + f(x)\lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \qquad \textit{(lim. rul. 3, 1)}$$

$$= f'(x)g(x) + f(x)g'(x) \qquad \textit{(deriv. def.)}$$

Exercise 8. *The second rule, the sum rule, is also to verify in the same way as in Example 16. Verify this rule.*

We also need a list of basic functions and their derivatives that often form the constituents of a function.

- $\dfrac{d}{dx} c = 0$ (the derivative of a constant is zero)

- $\dfrac{d}{dx} x^r = rx^{r-1}, \quad r \in \mathbb{R}$ (power rule)

- $\dfrac{d}{dx} x = x^0 = 1$ (special case of power rule)

- $\dfrac{d}{dx} \dfrac{1}{x} = -\dfrac{1}{x^2}$ (special case of x^r)

- $\dfrac{d}{dx} \sqrt{x} = \dfrac{1}{2\sqrt{x}}$ (also special case of x^r)

- $\dfrac{d}{dx} g(x)^r = r\, g(x)^{r-1} \dfrac{d}{dx} g(x)$ (general power rule)

2.1. Exercises. This section contains exercises on applying differentiation rules. It's important in the exercises that at each step in you calculation you *consciously* justify the step with the appropriate rule, especially in the beginning. Try to consider each application of a rule as suspicious.

Example 18. *Let f be $f(x) = ax + b$. Find $f'(x)$.*

ANSWER.
$$
\begin{aligned}
f'(x) &= \frac{d}{dx}(ax + b) \\
&= \frac{d}{dx}(ax) + \frac{d}{dx}b && \text{(sum rule)} \\
&= a \cdot \frac{d}{dx}x + 0 && \text{(constant multiple \& constant rule)} \\
&= a \cdot 1 && \text{(power rule)} \\
&= a
\end{aligned}
$$
□

Example 19. *Let f be $f(x) = x^4 - \frac{3}{x^2}$. Find $f'(x)$.*

ANSWER. Rewrite $f(x) = x^4 - \frac{3}{x^2} = x^4 - 3x^{-2}$.
$$
\begin{aligned}
f'(x) &= \frac{d}{dx}(x^4 - 3x^{-2}) \\
&= \frac{d}{dx}x^4 - \frac{d}{dx}3x^{-2} && \text{(sum rule)} \\
&= 4x^3 - 3\frac{d}{dx}x^{-2} && \text{(power \& constant multiple rule)} \\
&= 4x^3 - 3 \cdot (-2)x^{-3} && \text{(power rule)} \\
&= 4x^3 + 6x^{-3}
\end{aligned}
$$
□

Example 20. *Let f be $f(t) = at^2 + st + s^2$. Find $f'(t) = \frac{d}{dt} f(t)$.*

ANSWER.

$$
\begin{aligned}
f'(t) &= [at^2 + st + s^2]' \\
&= [at^2]' + [st]' + [s^2]' && \text{(sum rule)} \\
&= 2at + s + 0 && \text{(power \& constant rule)} \\
&= 2at + s
\end{aligned}
$$

Example 21. *Let f be $f(x) = \sqrt{1 - x^2}$. Find $f'(x)$.*

ANSWER. Rewrite $f(x) = \sqrt{1 - x^2} = (1 - x^2)^{1/2}$, and apply the power and differentation rules:

$$
\begin{aligned}
f'(x) &= \frac{d}{dx}(1 - x^2)^{1/2} && \text{(recognize structure $g(x)^r$)} \\
&= \frac{1}{2}(1 - x^2)^{-1/2}\frac{d}{dx}(1 - x^2) && \text{(general power rule)} \\
&= \frac{1}{2\sqrt{1 - x^2}}(\frac{d}{dx}1 - \frac{d}{dx}x^2) && \text{(sum rule)} \\
&= \frac{1}{2\sqrt{1 - x^2}}(0 - 2x) \\
&= \frac{-x}{\sqrt{1 - x^2}}
\end{aligned}
$$

$\square$

Example 22. *Let f be $f(t) = t + \dfrac{1}{t}$. Find $f''(t) = \frac{d^2}{dt^2} f(t)$.*

ANSWER.

$$
f'(t) = [t + t^{-1}]' = [t]' + [t^{-1}]' = 1 + (-1)t^{-2} = 1 - t^{-2}
$$

Hence,

$$
f''(t) = [f'(t)]' = [1 - t^{-2}]' = [1]' - [t^{-2}]' = 0 - (-2)t^{-3} = 2t^{-3},
$$

so that $f''(t) = 2t^{-3}$.

$\square$

Example 23. *Let f be $f(x) = (2x^3 - 5x)(3x + 1)$. Find $f'(x)$.*

ANSWER.

$$
\begin{aligned}
f'(t) &= [(2x^3 - 5x)(3x + 1)]' \\
&= (2x^3 - 5x)'(3x + 1) + (2x^3 - 5x)(3x + 1)' && \text{(product rule)} \\
&= ([2x^3]' - [5x]')(3x + 1) + (2x^3 - 5x)([3x]' + [1]') && \text{(sum rule)} \\
&= (2[x^3]' - 5[x]')(3x + 1) + (2x^3 - 5x)(3[x]' + 0) && \text{(constant multiple rule)} \\
&= (2[3x^2] - 5[1])(3x + 1) + (2x^3 - 5x)(3[1] + 0) && \text{(power rule)} \\
&= (6x^2 - 5)(3x + 1) + 3(2x^3 - 5x)
\end{aligned}
$$

Example 24. *Let h be $h(x) = \dfrac{x^2 - 4 + x}{(x+1)(x-1) + x}$. Find $h'(x)$.*

ANSWER. Rewrite

$$h(x) = \frac{x^2 - 4 + x}{(x+1)(x-1) + x} = \frac{x^2 - 1 + x - 3}{x^2 - 1 + x},$$

and define

$$g(y) = x^2 - 1 + x, \quad \text{and} \quad f(y) = \frac{y - 3}{y}.$$

Then $h(x) = f(g(x))$. Using the chain rule,

$$h'(x) = [f(g(x))]' = f'(g(x))\, g'(x).$$

For $f'(y)$ we have

$$f'(y) = \frac{d}{dy} \frac{y - 3}{y} = [1 - \frac{3}{y}]' = 0 - 3[y^{-1}]' = -3(-1)y^{-2} = \frac{3}{y^2}.$$

For $g'(x)$ we have

$$g'(x) = [x^2 - 1 + x]' = 2x + 1.$$

Hence, the chain rule gives $h'(x) = \dfrac{3}{(x^2 - 1 + x)^2} \cdot (2x + 1) = \dfrac{6x + 3}{(x^2 - 1 + x)^2}$. The answer could of course also have been found with the quotient rule.   □

Exercise 9. *Find the derivatives of the indicated functions.*

(1) $\sqrt{x+1} - \sqrt{x}$

(2) $\dfrac{x^2 - 2\sqrt{x}}{x}$

(3) $\sqrt{x} + 1/\sqrt[3]{x^4}$

(4) $\dfrac{(x^2 + 1)^2 - x^4}{x^2 + 1}$

(5) $\dfrac{x^2 - 5x}{x^2 - 3x - 10}$

Exercise 10. *Find the equation $y = ax + b$ of the tangent of the graph of the function $f(x) = 1 - x^2$ in the point $(-1, 0)$.*

3. Applications

Function properties. The derivative of a function gives information about the local behavior of the function. In discussing secant lines, we noticed that

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

is the slope of the secant line approximation of the function. The limit of this slope as $b \to a$ (if it exists) is the derivative of $f(a)$ at a, and the slope of the tangent line to the curve of $f(x)$. A positive slope indicates that the tangent is *increasing*, and hence that $f(x)$ is also increasing. A negative slope indicates that the tangent is *decreasing* and hence that $f(x)$ is also decreasing. A mathematically rigorous way of saying that a function is increasing is to say that "$f(b) > f(a)$ if and only if $b > a$". Similarly, to say that a function is decreasing in a mathematically rigorous

way is that "$f(b) > f(a)$ if and only if $a > b$". One even makes the distinction between strongly increasing functions ($f(b) > f(a)$ if and only if $b > a$), and weakly increasing functions ($f(b) \geq f(a)$ if and only if $b > a$). Such functions are sometimes referred to as *monotone*. What happens when $f'(x) = 0$? Then the slope is neither negative nor positive and so the tangent is neither increasing nor decreasing. The tangent is therefore horizontal, indicating that $f(x)$ doesn't change. This is illustrated in figure 2
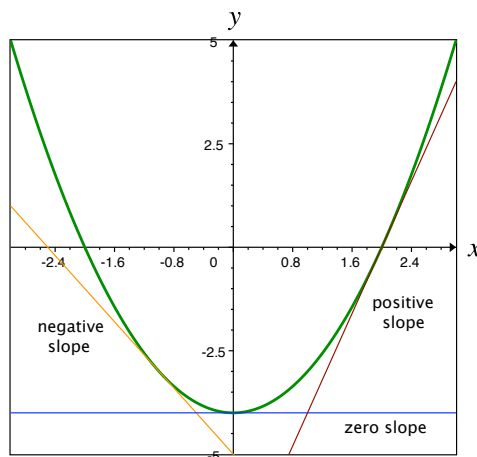

FIGURE 2. Tangents to the graph of the function $f(x) = x^2 - 4$. At different points the tangent is either positive, negative or zero, locally indicating the direction of change of the function for a small increase in x.

The second order derivative of $f(x)$, $f''(x)$, gives the change in the slope of the tangents of $f(x)$. If $f''(x)$ is positive, $f'(x)$ increases. If the the tangent lines to $f(x)$ are increasing, i.e., $f'(x) > 0$, the tangents get steeper and steeper. If the tangents have a negative slope, i.e., $f'(x) < 0$, then the tangent lines to $f(x)$ get less steep. In either case, the curve of $f(x)$ is changing its course to become greater—it is *bending upward*. In a region where $f''(x) > 0$ and hence, $f(x)$ is bending upward, $f(x)$ is said to be *convex*. If $f''(x)$ is negative, the opposite occurs: The slopes of the tangents get smaller, and hence the become more steep if they are decreasing, while they become less steep if they are increasing. In either case, $f(x)$ is changing its course to become smaller—it is *bending downwards*. In a region where $f''(x) < 0$ and hence $f(x)$ is bending downwards, $f(x)$ is said to be *concave*. The archetype of convex and concave functions is the parabola:

Example 25. *Let $f(x) = x^2$, and let $g(x) = -x^2$. The tangent lines of $f(x)$ have slope $f'(x) = [x^2]' = 2x$. Hence, for $x < 0$, the slopes of the tangents $f'(x)$ are negative, and so $f(x)$ is decreasing. For $x > 0$, the slopes $f'(x)$ are positive, and so $f(x)$ is increasing. At any point $f''(x) = [x^2]'' = [2x]' = 2 > 0$ for all x that we might want to consider, and hence, $f(x)$ is bending upwards everywhere. It is convex.*

Exercise 11. *Investigate the function $g(x)$ of Example 25 in the same way as it was done in that example for $f(x)$.*

Rate of change. The slope of the tangent of $f(x)$ at a point a locally gives the rate at which $f(x)$ is changing when x increases. Hence, at any point $f'(x)$ tells us how fast $f(x)$ changes. Similarly $f''(x)$ tells us how fast $f'(x)$ is changing. Hence, $f''(x)$ tells us how fast *the change in* $f(x)$ changes.

Example 26. *A sales person travels in her private airplane to a client. As a function of time (in minutes), the distance (in kilometers) she has traveled is given by*

$$s(t) = \frac{100t^2}{500 + t^2} + 2t.$$

What is here average speed after 50 minutes? What is here speed at $t = 25$? How fast changes her speed at $t = 25$—i.e., what is her acceleration at $t = 25$?

ANSWER. After 50 minutes, she has traveled $s(50) = \frac{250}{3} + 100 \approx 183$km. It has taken her 5/6-th of an hour, and so her average speed is approximately $183/(5/6) = 219\frac{6}{10}$ km/h. Her speed at $t = 25$ is the rate of change of her distance at $t = 25$. This is given by the derivative of s at $t = 25$,

$$\begin{aligned}
s'(t) &= \frac{[100t^2]'(500 + t^2) - 100t^2[500 + t^2]'}{(500 + t^2)^2} + [2t]' \\
&= \frac{100 \cdot 2t^1(500 + t^2) - 100t^2(2t^1)}{(500 + t^2)^2} + 2 \\
&= \frac{200t(500 + t^2) - 200t^3}{(500 + t^2)^2} + 2 \\
&= \frac{200t}{500 + t^2} - \frac{200t^3}{(500 + t^2)^2} + 2.
\end{aligned}$$

The speed at $t = 25$ is therefore $s'(25) \approx 3.98$ km/min, or about 238.5 km/h. Her acceleration is given by the change in speed, and hence is given by $s''(t)$.   □

Exercise 12. *Compute the acceleration at $t = 25$ in Example 26.*

Optimization. A function $f(x)$ has a local maximum (minimum) in a point a if $f(a) \geq f(x)$ ($f(a) \leq f(x)$) for all x close enough to a. If $f(x)$ is continuous at this point, the tangent line approximation should be horizontal at this point, because the function increases (decreases) on the left of this point, and decreases (increases) on the right of this point. So the slop of the tangent should change sign at this point, and so the slope should be exactly zero at the location of the maximum (minimum). Since the slope is given by the derivative, the derivative has to be zero at this point.[1] Therefore, we can find maxima and minima of a function by finding those values of x at which

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

That is, we have to solve this equation for x.

This equation is a necessary condition for a maximum or minimum of a continuously differentiable function. It is not sufficient however. There are points at which $f'(x) = 0$, which correspond neither to a maximum nor a minimum. Sufficient conditions for these extrema are

- $f(x)$ has a (local) minimum at a if $f'(a) = 0$ and $f''(a) > 0$, and
- $f(x)$ has a (local) maximum at a if $f'(a) = 0$ and $f''(a) < 0$.

If $f'(a)$ and $f''(a) = 0$, then $f(x)$ may have a maximum, minimum, or neither.

Example 27. *If $f(x) = x^2$, then $f'(0) = 0$, and $f''(0) = 2 > 0$; $f(x)$ therefore has a (local) minimum at 0.*

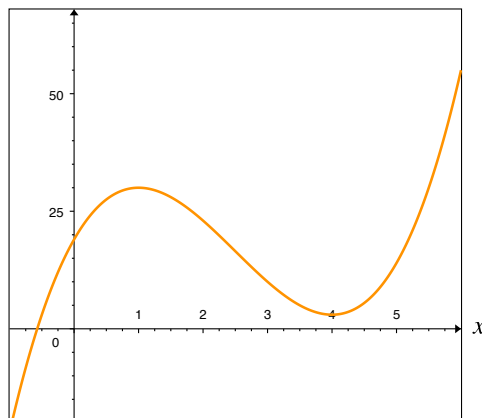[1] Strictly this is only true if the derivative is continuous in this point.

FIGURE 3. Extrema of the function in Example 29.

Example 28. *If $f(x) = -x^2$, then $f'(0) = 0$, and $f''(0) = -2 < 0$; $f(x)$ therefore has a (local) maximum at 0.*

Example 29. *If $f(x) = 2x^3 - 15x^2 + 24x + 19$, then $f'(x) = 6x^2 - 30x + 24$. To find the points where the slope of the tangent line is zero, we have to solve $f'(x) = 0$ or*

$$6x^2 - 30 + 24 = 6(x^2 - 5x + 4) = 0.$$

This is solved for

$$x = \frac{-(-5) \pm \sqrt{(-5)^2 - 4 \cdot 4}}{2}, \quad \text{that is, by } x = 4 \text{ and } x = 1.$$

Hence, $f(x)$ has a maximum or minimum at $x = 4$ and $x = 1$. The second derivative is $f''(x) = 12x - 30$, and $f''(1) = -18 < 0$ so that $f(x)$ has a (local) maximum at $x = 1$, and $f''(4) = 18 > 0$ so that $f(x)$ has a (local) minimum at $x = 4$. The function is graphed in figure 3.

Example 30. *If $f(x) = x^3$, then $f'(0) = 0$, and $f''(0) = 0$; whether $f(x)$ has a (local) extremum is therefore undecided. The graph of the function shows that it has no extremum at 0.*

Example 31. *If $f(x) = x^4$, then $f'(0) = 0$, and $f''(0) = 0$; whether $f(x)$ has a (local) extremum is therefore undecided. The graph of the function shows that it has a minimum at 0.*

Example 32. *If $f(x) = |x|$, then $f'(x)$ does not exist at 0, and therefore is not continuously differentiable around 0 and not differentiable at 0, yet f has a local minimum at 0.*

Many problems in science, engineering, and business can be brought down to finding maxima or minima (collectively called *extrema*) of functions.

Exercises

The exercises in this section ask you to find maxima or minima of functions.

Example 33. Maximizing volume *Parcel post regulations require that packages must have a length plus girth of no more than 2 meters. Design a cylindrical package that maximizes the volume while satisfying the regulatory constraints.*

SOLUTION. For a cylinder the girth of the package equals $2\pi r$, where r is the radius of the circular end. Volume of a cylinder is give by $\pi r^2 l$, where l is the length of the cylinder. To take maximal advantage of the regulations we make sure that the girth plus length equals exactly the maximum allowed. Hence, we have the following equations,

$$V = \pi r^2 l \qquad \text{(objective function)}$$
$$2\pi r + l = 2 \qquad \text{(constraint equation)}$$

The first equation specifies the quantity that we wish to maximize by varying r and/or l. The second equation imposes a constraining relation between r. We solve the constraint equation for l and substitute the expression into the objective function: $l = 2 - 2\pi r$, hence, we have

$$V = \pi r^2 (2 - 2\pi r).$$

Let $f(r) = \pi r^2 (2 - 2\pi r)$. We would like to find r where $f(r)$ attains a maximum. The function is plotted in the figure below.
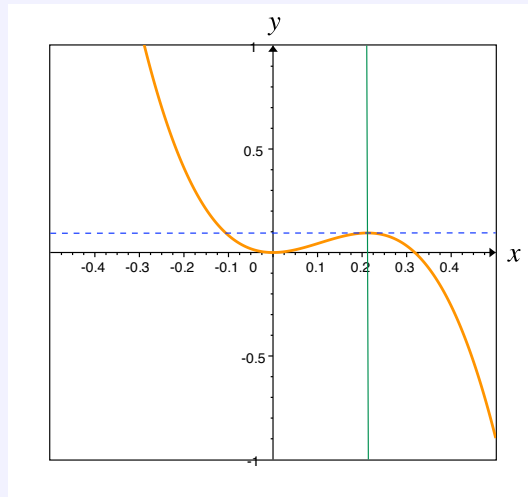
First, we set the derivative equal to zero and solve for r,

$$f'(r) = 4\pi r - 6\pi^2 r^2 = 0 \iff r(1 - \frac{3}{2}\pi r) = 0 \iff r = 0 \text{ or } r = \frac{2}{3\pi}$$

Although it is clear that the solution $r = 0$ is not of much use, next we check if the alternate solution is a local maximum:

$$f''(r) = 4\pi - 12\pi^2 r \implies f''(2/3\pi) = 4\pi - 12\pi^2 (2/3\pi) = -4\pi < 0,$$

therefore $f(r)$ has a maximum at $r = 2/3\pi$.

Hence, the volume is maximized, subject to the constraint, when $r = 2/3\pi$. □

Exercise 13. *A large (cylindric) soup can is to be designed so that the can will hold 16π cubic inches (about 28 ounces) of soup. Find the radius r and height h for which the amount of metal needed is as small as possible.*

Chapter 2

Transcendental functions: exp, ln, sin, cos, tan

In this chapter we deal with special functions that are encountered widely throughout applied and theoretical analysis.

1. The exponential function

The exponential function can be motivated by differentiation of general exponential functions a^x; Recall the definition of the derivative of a function in a point x,

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

and let us apply it to $f(x) = a^x$:

$$[a^x]' = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} = \lim_{h \to 0} \frac{a^x a^h - a^x}{h} = \lim_{h \to 0} a^x \frac{a^h - 1}{h} = a^x \lim_{h \to 0} \frac{a^h - 1}{h} \quad \text{(limit rule 1)}.$$

That is, $[a^x]'$ is equal to itself times the value

$$\lim_{h \to 0} \frac{a^h - 1}{h}.$$

Note that this value does not depend on x. This limit is determined by the value of the function a^x itself in the neighborhood of $x = 0$. Take a look at figure 1.

We notice a couple of things

- All exponential functions seem to go through the point $(0, 1)$, and this is indeed true, as $a^0 = 1$ for all $a > 0$.
- All exponential function seem to be either strictly increasing or strictly decreasing, except for $a = 1$, in which case $a^x = 1^x = 1$ is a constant function. Clearly this is true, since $a^{x+1} = (a^x)a > a^x$ iff $a > 1$ (hence a^x is increasing), and $a^{x+1} = (a^x)a < a^x$ if $a < 1$ (hence a^x is decreasing).
- The slopes of the tangents in the point $x = 0$ are all different and depend on the value of a.

The number e is now defined to be that number for which

$$\lim_{h \to 0} \frac{e^h - 1}{h} = 1;$$

that is, it is that base number such that the derivative of the function $f(x) = e^x$ evaluated in $x = 0$ equals 1: $f'(0) = 1$. But we just found

$$[a^x]' = a^x \lim_{h \to 0} \frac{a^h - 1}{h},$$

and so, for $a = e$,

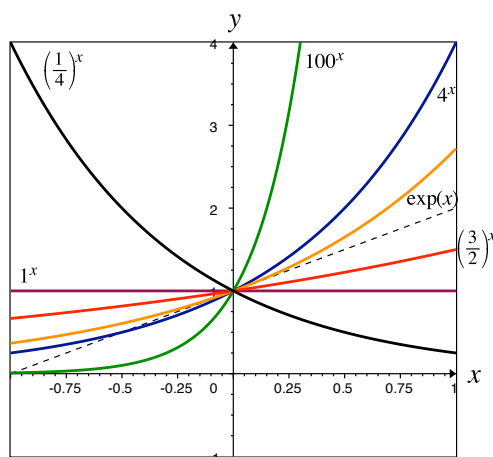$$[e^x]' = e^x \lim_{h \to 0} \frac{e^h - 1}{h} = e^x \cdot 1 = e^x.$$

FIGURE 1. Exponential functions, for different bases. The dashed line is the tangent to the exponential function with base e.

The function e^x is the only function whose derivative is equal to itself.

Because of this property, and because

$$a^x = (a)^x = (e^{\log_e a})^x = e^{x(\log_e a)},$$

for any $a > 0$—that is, an exponential function can always be re-expressed in terms of a power of the number e—it is much easier to work with e as a base than any other number a. This is why the function e^x is often (colloquially) referred to as *the* exponential function, and why it has the special symbol

$$\exp(x) = e^x.$$

The decimal expansion of e is approximately

$$e = 2.7182818284590452353602874713526624977572470936999595\ldots,$$

and it can be proven that e is irrational (*irrational?*).

Clearly the same rule apply for e^x as for any other exponential function. To recapitulate

$$\exp(0) = e^0 = 1, \quad \exp(1) = e^1 = e,$$
$$\exp(x + y) = e^{x+y} = e^x e^y = \exp(x)\exp(y), \quad \exp(x)^y = (e^x)^y = e^{xy} = \exp(xy).$$

Most people find the e^x notation easier to work with.

2. The natural logarithm

The inverse function of e^x, $\log_e(x)$ is called the *natural logarithm* and is commonly denoted $\ln(x)$. Clearly the same rules apply to ln as to $\log_a$:

$$\ln(1) = 0, \quad \ln(e) = 1, \quad \ln(xy) = \ln(x) + \ln(y), \quad \text{and } \ln(x^y) = y\ln(x).$$

Figure 2 displays both $\exp(x)$ and $\ln(x)$. Note that they are mirror images of each other in the diagonal line (the line $y = x$), as all inverse function pairs are. Also note that $\exp(x)$ never becomes negative (like all exponential functions), and that $\ln(x)$ is not defined for $x < 0$. In the
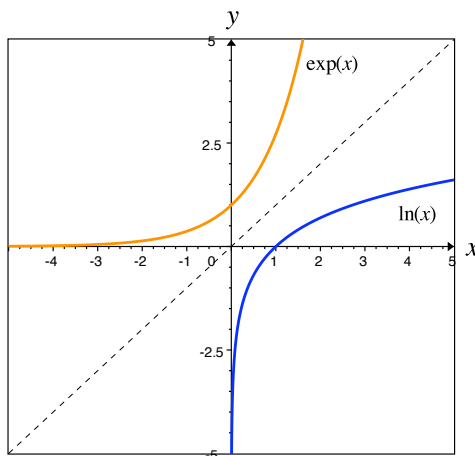
FIGURE 2. The exponential function $\exp(x)$ and its inverse function $\ln(x)$. They are mirror images from each other in the line $y = x$ (dashed line).

limit $x \to \infty$, $\exp(x) \to \infty$ and $\ln(x) \to \infty$. In the limit $x \to -\infty$, $\exp(x) \to 0$. In the limit $x \to 0$, $\ln(x) \to -\infty$. More importantly, $\exp(x)$ grows very fast, while $\ln(x)$ grows very slowly.

3. Derivatives

We have already seen that the derivative of $\exp(x)$ is $\exp'(x) = \exp(x)$, i.e., itself. Now, lets consider the derivative of e^{cx}: By the chain rule for differentiation, $[f(g(x))]' = f'(g(x))\, g'(x)$,

$$[\exp(cx)]' = \exp'(cx)[cx]' = \exp(cx)c[x]' = c\exp(cx),$$

or, in terms of the number e,

$$[e^{cx}]' = ce^{cx}.$$

This allows us to compute the derivative of a^x, because, as before, $a^x = e^{x \log_e a}$.

Exercise 14. *Compute $[a^x]'$ using $a^x = e^{x \log_e a}$.*

Note that we now have circumvented the need to explicitly compute the limit

$$\lim_{h \to 0} \frac{a^h - 1}{h}.$$

In fact, it shows that $\lim_{h \to 0} \frac{a^h - 1}{h} = \ln a$.

The chain rule gives a far more general rule: $[\exp(g(x))]' = \exp'(g(x))\, g'(x) = \exp(g(x))\, g'(x)$. In terms of e:

$$[e^{g(x)}]' = e^{g(x)} g'(x) \qquad \text{or} \qquad \frac{d}{dx} e^{g(x)} = e^{g(x)} \frac{d}{dx} g(x).$$

This also gives us the derivative of $\ln(x)$:

$$1 = [x]' = [e^{\log_e x}]' = [e^{\ln(x)}]' = e^{\ln(x)} \ln'(x) = x \ln'(x),$$

and consequently

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

Example 34. *The derivatives of* exp *and* ln *allow us to verify the power rule of the previous chapter. Recall that according to the power rule*

$$[x^r]' = rx^{r-1}.$$

Verify this rule.

SOLUTION. The first part is easy, the second part is more tricky:

Suppose $x > 0$. Then we can always rewrite $x^r = e^{\ln x^r}$, and

$$(9) \qquad\qquad [x^r]' = [e^{\ln x^r}]' = e^{\ln x^r}[r \ln x]' = x^r \left(r \frac{1}{x} \right) = r\, x^{r-1}.$$

We now need to show that it is also true if $x < 0$.

Suppose $x < 0$, then r must be an integer for x^r to make sense, and we can write $x^r = (-1)^r |x|^r$, and therefore,

$$[x^r]' = [(-1)^r |x|^r]' = (-1)^r [|x|^r]',$$

by the constant multiple rule for differentiation.

Since $|x|$ is positive, equation (9) together with the chain rule imply that $[|x|^r]' = r|x|^{r-1} \frac{d}{dx} |x|$. Now, since we assumed $x < 0$, $|x| = -x$, and $\frac{d}{dx} |x| = \frac{d}{dx}(-x) = -1$. Therefore,

$$[(-1)^r |x|^r]' = (-1)^r\, r|x|^{r-1}(-1) = (-1)^2 rx^{r-1} = rx^{r-1}.$$

$\square$

4. Applications

Exponential growth and decay. The relationship $[e^{cx}]' = ce^{cx}$ is more important than it may seem: If we define $y = e^{cx}$, then we can state it as

$$\frac{dy}{dx} = cy,$$

essentially stating that "the rate of change in y is proportional to the value of y". This equation has many applications throughout the sciences. It is an elementary example of a *differential equation.*

Example 35. *Bacteria divide at a constant rate. That is, the change in the number of bacteria at time t is proportional to the number of bacteria at time t. Stated mathematically, if $y(t)$ denotes the number of bacteria at time t, then*

$$\frac{dy}{dt} = ky.$$

Find a function $y(t)$ that satisfies this equation.

SOLUTION. From the discussion of e^{cx}, we guess that e^{kt} satisfies this equation. To verify this, we use $y = e^{kt}$ as a trial solution (*ansatz*), and we differentiate: $\frac{d}{dt} y = \frac{d}{dt} e^{kt} = ke^{kt} = ky$. Therefore, the function $y(t) = e^{kt}$ is called a *solution* to the equation. *A* solution, because $y = Ce^{kt}$ is also a solution for arbitrary C—that is, e^{kt} is not the *only* solution! $\square$

Logarithmic differentiation. The exponential and natural logarithm functions can also be quite useful in reducing the amount of work when calculating derivatives of functions that are constituted by *products of a large number of constituent functions*: Because of the chain rule,

$$\frac{d}{dx} \ln f(x) = \frac{1}{f(x)} f'(x) = \frac{f'(x)}{f(x)}.$$

Therefore

$$f'(x) = f(x) \frac{d}{dx} \ln f(x).$$

So how does this help? The logarithm turns the product into a sum, because $\ln(xyz) = \ln x + \ln y + \ln z$. The sum rule of differentiation is much easier than the product rule. This technique is called *logarithmic differentation*

Example 36. *Find the derivative of $f(x) = (x^2 + 1)(x^3 - 3)(2x + 5)$.*

ANSWER. Using logarithmic differentation
$$[\ln f(x)]' = [\ln(x^2 + 1) + \ln(x^3 - 3) + \ln(2x + 5)]'$$
$$= [\ln(x^2 + 1)]' + [\ln(x^3 - 3)]' + [\ln(2x + 5)]'$$
$$= \frac{1}{x^2 + 1}[x^2 + 1]' + \frac{1}{x^3 - 3}[x^3 - 3]' + \frac{1}{2x + 5}[2x + 5]'$$
$$= \frac{2x}{x^2 + 1} + \frac{3x^2}{x^3 - 3} + \frac{2}{2x + 5},$$
and hence,
$$f'(x) = f(x)\frac{d}{dx}\ln f(x) = (x^2 + 1)(x^3 - 3)(2x + 5)\left(\frac{2x}{x^2 + 1} + \frac{3x^2}{x^3 - 3} + \frac{2}{2x + 5}\right).$$

□

Example 37. *Determine the derivative of*
$$f(x) = x^x.$$

ANSWER. Using logarithmic differentation
$$\frac{d}{dx}\ln f(x) = [x\ln x]' = \ln(x) + x\frac{1}{x} = \ln(x) + 1$$
and hence,
$$f'(x) = f(x)\frac{d}{dx}\ln f(x) = x^x(\ln(x) + 1).$$

□

Example 38. *Find the derivative of*
$$f(x) = \frac{\sqrt[3]{x + 1}}{(x + 2)\sqrt{x + 3}}.$$

ANSWER. Using logarithmic differentiation
$$\frac{d}{dx}\ln f(x) = [\ln(x + 1)^{1/3} + \ln(x + 2)^{-1} + \ln(x + 3)^{-1/2}]'$$
$$= \frac{1}{3}\frac{[x + 1]'}{x + 1} - \frac{[x + 2]'}{x + 2} - \frac{1}{2}\frac{[x + 3]'}{x + 3}$$
$$= \frac{1}{3(x + 1)} - \frac{1}{x + 2} - \frac{1}{2(x + 3)}$$
and hence,
$$f'(x) = f(x)\frac{d}{dx}\ln f(x) = \frac{\sqrt[3]{x + 1}}{(x + 2)\sqrt{x + 3}}\left(\frac{1}{3(x + 1)} - \frac{1}{x + 2} - \frac{1}{2(x + 3)}\right).$$

□

Derivatives of inverse functions. The technique we used to find the derivative of $\ln(x)$ can be used more generally to find derivatives of inverse functions.

Let $f(x)$ be a one-to-one function, and let $g(y)$ be its inverse (g may also be written f^{-1} but that will clutter the notation). Then we know that

$$g(f(x)) = x,$$

by definition. Suppose we know the derivative $f'(x)$ of $f(x)$. Let us use the chain rule to derive the derivative of $g(y)$:

$$1 = [x]' = [g(f(x))]' = g'(f(x))f'(x).$$

Therefore

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

but $y = f(x)$ and $x = g(y)$, hence,

$$g'(y) = \frac{1}{f'(g(y))}, \quad \text{or} \quad \frac{d}{dy}f^{-1}(y) = \frac{1}{f'(f^{-1}(y))}, \quad \text{or} \quad \frac{dg}{dy} = \frac{1}{\frac{df}{dx}|_{g(y)}} \quad \text{or} \quad \frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}.$$

Example 39. *Let $f(x) = \frac{x}{1-x}$, then $y = x/(1-x) \iff y - yx = x \iff y = (1+y)x \iff x = y/(y+1)$, thus $g(y) = f^{-1}(y) = \frac{y}{y+1}$. Verify the inverse function rule.*

ANSWER. The derivative of f is:

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

Applying the inverse function rule to $f(x)$,

$$g'(y) = \frac{1}{\frac{1}{(1-g(y))^2}} = (1-g(y))^2 = \left(1 - \frac{y}{y+1}\right)^2 = \left(\frac{1}{y+1}\right)^2.$$

Direct differentiation of $g(y)$ indeed yields $g'(y) = \frac{1}{(y+1)^2}$.   □

The inverse function rule has more use in cases where no closed form expression for the inverse can be found. A more prominent example will be postponed to the next section.

5. Trigonometric functions

Trigonometric function relates the length of the sides of a triangle to the angles the sides make. The sine of the indicated angle α in figure 3 is defined as the ratio between the length of the opposite side, y, and the length of the hypothenuse (slant side), r—i.e., $\sin(\alpha) = y/r$. The cosine of the angle is defined as the ration between the length adjacent side, x, and the hypothenuse—i.e., $\cos(\alpha) = x/r$. The tangent of the angle is defined as the slope of the hypothenuse, the ratio between length of the opposite side and the adjacent side—i.e., $\tan(\alpha) = y/x$. To memorize this, the word $SOH - CAH - TOA$ is often recommended.[1]

Some properties are

$$\sin(-\alpha) = -\sin(\alpha)$$
$$\cos(-\alpha) = \cos(\alpha)$$
$$\tan(\alpha) = \frac{\sin(\alpha)}{\cos(\alpha)}$$
$$\sin(\alpha)^2 + \cos(\alpha)^2 = 1.$$

---

[1] In Dutch: *soscastoa*—sinus is overstaande gedeelde door schuine, cosinus is overstaande gedeeld door schuine, tangens is overstaande gedeeld door aanliggende.
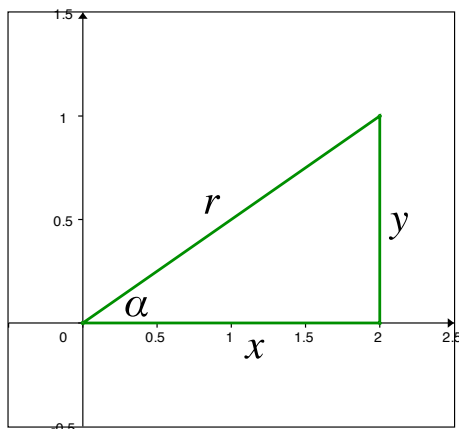
FIGURE 3. Triangle.

The last equation is of course the Pythagorean theorem. Despite inconsistencies with other uses of the notation $f^2(x)$, quite often $\sin^2(\alpha)$ is used to indicate the square for functions such as $\sin, \cos, \tan, \sec$, etc.

Figure 4 shows how sin and cos oscillate as a function of α, when the sides make up a triangle bounded by the unit circle (the hypothenuse always of unit length).
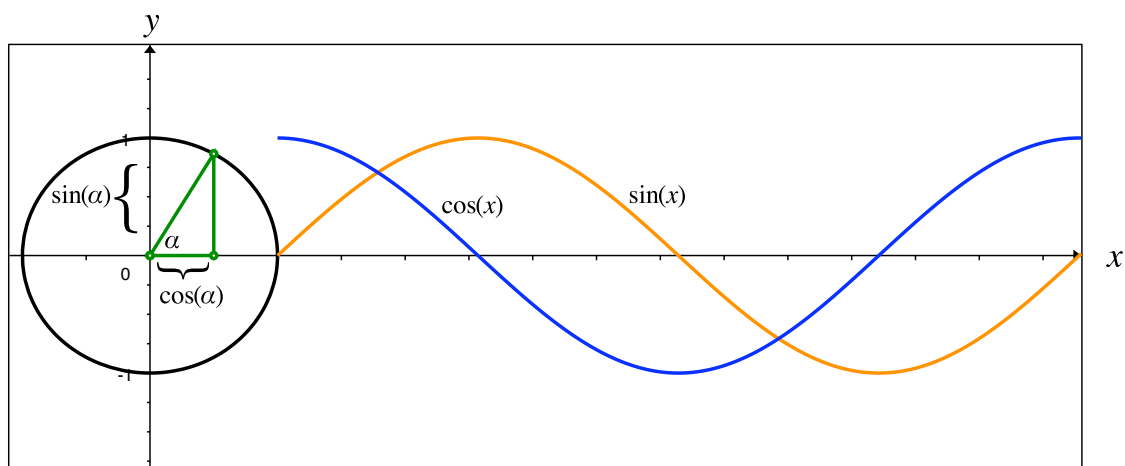
FIGURE 4. Values of sin and cos as the angle revolves the hypothenuse around the unit circle anti-clockwise. The sin is read of from the vertical axis; the cos is read off from the horizontal axis within the unit circle. These values are plotted as a function of α in orange (sin) and blue (cos).

Derivatives of sin, cos, and tan. Without proof, we give the derivatives of sin, cos, and tan, along with the derivative rules discussed so far.

TABLE 1. Derivatives of transcendental functions

$f(x)$	$\exp(x)$	$\ln(x)$	$\sin(x)$	$\cos(x)$	$\tan(x)$
$f'(x)$	$\exp(x)$	$\dfrac{1}{x}$	$\cos(x)$	$-\sin(x)$	$\dfrac{1}{\cos(x)^2}$

Example 40. $\sin(x)$ *is monotonously decreasing on the interval* $[-\pi/2, \pi/2]$. *On that interval we can therefore define an inverse for* $\sin(x)$. *This inverse is often denoted* $\arcsin(x)$ *or* $\sin^{-1}(x)$. *Find the derivative of* $\arcsin(x)$.

ANSWER. We can use the inverse function rule for derivatives discussed earlier. According to that rule if $g(y)$ is the inverse of $f(x)$ with known derivative $f'(x)$, then

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

In this case, $f(x) = \sin(x)$, and $f'(x) = \cos(x)$ (from table 1), and $g(x) = \arcsin(x)$. Then, using $\sin(x)^2 + \cos(x)^2 = 1$,

$$\frac{d}{dy} \arcsin(y) = \frac{1}{\cos(\arcsin(y))} = \frac{1}{\sqrt{1 - \sin(\arcsin(y))^2}} = \frac{1}{\sqrt{1 - y^2}}.$$

□

Example 41. *Compute* $\dfrac{d}{dx} \sin(x^2)$.

ANSWER.

$$\frac{d}{dx} \sin(x^2) = \sin'(x^2)[x^2]' = \cos(x^2) \cdot 2x.$$

□

Example 42. *Compute* $\dfrac{d}{dx} \sin(x)^2$.

ANSWER.

$$\frac{d}{dx} \sin(x)^2 = 2\sin(x)[\sin(x)]' = 2\sin(x)\cos(x).$$

□

Example 43. *Compute* $\dfrac{d}{dx} \ln\cos(3x)$.

ANSWER.

$$[\ln\cos 3x]' = \ln'(\cos(3x))[\cos(3x)]' = \frac{1}{\cos(3x)} \cos'(3x)[3x]' = \frac{1}{\cos(3x)}(-1)\sin(3x) \cdot 3$$

$$= -3 \cdot \frac{\sin(3x)}{\cos(3x)} = -3\tan(3x).$$

□

Note that because $[\sin(x)]' = \cos(x)$, and $[\cos(x)]' = -\sin(x)$, we have

$$[\sin(x)]'' = [\cos(x)]' = -\sin(x),$$

that is, the second order derivative of $\sin(x)$ is itself multiplied by -1. Hence, similar to the equation $y' = y$ for $y = e^x$, we have

$$y'' = -y$$

for $y = \sin(x)$. The same equation holds for $y = \cos(x)$. Furthermore, $[\sin(cx)]' = \cos(cx)[cx]' = c\cos(cx)$, by the chain rule, and $[\sin(cx)]'' = -c^2 \sin(cx) = k\sin(cx)$, a constant times itself. Hence, $\sin(cx)$ "solves" the equation

$$y'' = ky,$$

another elementary differential equation—a second order linear differential equation (as opposed to $y' = cy$, which is a first order linear differential equation). By "solving", here we mean to say that if we start with the equation and we have to find y, we say that we have have solved the equation if we find a function y that satisfied it. The same equation is also "solved" by $\cos(cx)$. Like the equation $y' = cy$ it has many application throughout the sciences. Both for instance essential in electrical models for neurons. We will not discuss these equations further.

6. Exercises

This section contains various exercises on differentiation, using all the rules discussed so far.

Exercise 15. *Determine derivatives as indicated.*

(1) $f(x) = \exp(x^2 + x - 1)$. *Find* $f'(x)$.

(2) $f(x) = x \ln x$. *Find* $f'(x)$.

(3) $g(t) = (3 - t)^6$. *Find* $\dfrac{d}{dt} g(t)$.

(4) *Let* $h(u) = \dfrac{u}{u^2 + 1}$. *Compute* $h'(u)$.

(5) $f(x) = \ln(x + \sqrt{x^2 + a^2})$, *where* $a > 0$. *Compute* $f'(x)$.

(6) $f(x) = \ln|x|$. *Compute* $f'(x)$.

(7) *Determine* $\dfrac{d}{dx} (\sqrt{x} - 2 e^x)$.

(8) *Find* $[x e^{-1/x}]'$.

(9) $f(x) = \dfrac{x^2 - 2\sqrt{x}}{x}$. *Determine* $f'(x)$.

(10) *Show that* $\dfrac{d}{dx} \tan(x) = \dfrac{1}{\cos(x)^2}$.

(11) *Compute* $f'(x)$, *if* $f(x) = (4 + \sin x)^{(\cos x - \sin x)}$.

(12) $f(x) = \sqrt{x} + \dfrac{1}{\sqrt[3]{x^4}}$. *Determine* $f'(x)$.

(13) *Compute the derivative of* $\ln\left|\frac{x^2-4}{2x+5}\right|$.

(14) $f(x) = \sin(\tan(\sqrt{1 - x^3}))$. *Determine* $f'(x)$.

(15) $f(x) = x\cos(x)$. *Compute* $f''(x)$.

(16) $f(x) = \sin(x^2)$. *Compute* $f''(x)$.

(17) $f(x) = \dfrac{(x^2 + x + 2)(3x - 2)(x^4 - x^2 + 1)}{\sqrt[m]{x^3 + x - 1}}$. Find $f'(x)$. Hint: *use logarithmic differenti-*

   ation.

(18) *Let $f(x) = g(e^x)$. Express $f'(x)$ in terms of the derivative of $g(y)$.*

(19) *The function $f(x) = \dfrac{1}{1 + e^{-ax}}$ is used in psychometrics as an item response function,*

   and in connectionist psychology as activation function for neural networks. Show that

   $f(x)$ is (monotonously) increasing if $a > 0$.

(20) *Show that for $f(x) = \dfrac{1}{1 + e^{-x}}$, $f'(x) = f(x)(1 - f(x))$.*

Chapter 3

Integral calculus

In the last chapter we encountered the (elementary differential) equation

$$y' = y,$$

which often shows up in the sciences as soon as the change of a variable is related to is value (like the growth of a bacterial colony). This equation resulted from reasoning about the process under consideration (each bacterium as at any point in time a certain probability to divide into two bacteria, and so the change in the number of bacteria at any given moment is proportional to the amount of bacteria that exist at that time), and we asked which for which function y it would be true. The answer was $y = e^t$ (where t is used because we are talking about changes in time).

More generally, we can have a function $f(t)$, and the equation

$$y' = \frac{dy}{dx} = f(t),$$

and can ask the question "for which function $y(t)$ of t is this equation true?" From Chapter 0, you should be primed to recognize such a question as an *inversion*: So far, we have been asking "given a function $f(x)$, how do we determine $f'(x) = \frac{d}{dx} f(x)$?". In this chapter we will ask "given a function $f'(x)$, for which $f(x)$ is $\frac{d}{dx} f(x) = f'(x)$ true?"—that is, we will try to find ways to invert or "undo" the differentiation of a function.

You may object that this is not what integration is all about—at least not according to what you have learned or heard previously: that integration is about *calculating (surface) areas*. In essence, this is correct—the word "integral" means "whole" and is therefore related to summing parts. It turns out however, that Newton and Leibniz both independently and simultaneously discovered that the processes of calculating surface areas on the one hand, and "undoing differentiation" on the other, are one and the same thing, even though they are *conceptually* completely *different* things. This fact is known as "the Fundamental Theorem of Calculus", which we will talk about at the end of this chapter. After this chapter, we're essentially done with calculus. All that comes after is extra, or at least is called more properly 'analysis'.

Before turning to integration, we make one curious observation: Previously, when we introduced the concept of *inversions* and *inverse functions*, we were talking about "undoing" the operations that were carried out by a given function on a *number* to produced a new *number*. Differentiation however, is an operation that is carried out on functions, and produces a new *function* (which is tightly related to the original function as we shall see). Because differentiation always gives one and the same result function, one could say that "$\frac{d}{dx}$" is a 'function'! The difference between this 'function' and the functions that we have considered (i.e., the functions on which $\frac{d}{dx}$ operates) is that for this 'function', the domain is the collection of differentiable functions (often denoted C^1), and its range is in the collection of continuous functions (often denoted C^0). It is a mapping between functions, instead of a mapping between numbers. This observation progresses us into the realm of a field of mathematics known as 'linear algebra'.

1. Anti-differentiation

Inverting or "undoing" the effects of differentiation on a function is the opposite of differentiation, and, for want of a better word, is called *anti-differentiation*. Because its relation calculating surface areas, it is also termed *indefinite integration*.

As indicated anti-differentiation tries to find the function $F(x)$ such that when differentiated, it is equal to a given function $f(x)$. That is, it tries to find the function (solution) $F(x)$ to the equation

$$\frac{d}{dx} F(x) = f(x).$$

The function $F(x)$ is called the *primitive function*, or *primitive* for short, of $f(x)$. It is also denoted

$$\int f(x)dx,$$

for reasons that will become apparent throughout this chapter. In this notation, $f(x)$ is called the *integrand*. The symbol dx will be clarified in the next section, but it is always part of the notation to indicate the variable with respect to which the primitive of $f(x)$ is to be differentiated to regain $f(x)$. This is necessary for antiderivatives such as $\int (t^2 + s^3)\, ds$.

An immediate observation is that primitives are not unique: If $F(x)$ is a primitive of $f(x)$, then for any constant $C \in \mathbb{R}$, $F(x) + C$ is also a primitive of $f(x)$:

$$[F(x) + C]' = F'(x) + 0 = f(x).$$

Hence, primitives are said to be determined *up to an additive constant of integration*. This arbitrary additive constant is the reason for the adjective 'indefinite' in the term "indefinite integration."

Given a function $f(x)$, how do we find its primitive? The most fundamental technique of finding primitives is through recognition of the derivative. It is therefore imperative to have seen a lot of derivatives (that is, lots of practice with finding derivatives), and to be able to differentiate simple expressions mentally.

Example 44. *Find* $\displaystyle\int 2x dx$.

ANSWER. We recognize $2x$ as the derivative of x^2, so our solution is $F(x) = x^2$.
To verify that $F(x)$ is indeed a primitive of $f(x)$, we verify that it satisfies the equation $\frac{d}{dx} F(x) = f(x)$:

$$F'(x) = [x^2]' = 2x = f(x).$$

Notice however, that $F(x) = x^2 + C$, where $C \in \mathbb{R}$ is a constant, is also a solution. All solutions must be of this form (*why?*), and so it is the most general solution possible. When asked for an anti-derivative, it is understood that one *always* gives the most general solution. Hence, our answer is

$$F(x) = x^2 + C.$$

Not giving the most general solution, i.e., omitting the constant of integration is an error.   □

Exercise 16. *Show that $F(x) = x^2 + C$ in Example 44 is also a primitive of $f(x) = 2x$.*

As opposed to finding the derivative of a function—which is perhaps laborious and error prone, but can always be done given enough labour and effort—finding the primitive of a function is generally very difficult. Part of the difficulty lies in the fact that it is not always clear if the primitive can be expressed in terms arithmetic operations on elementary functions such as $\exp, \ln, \sin$.

Example 45. *For the function* $f(x) = \dfrac{1}{2\pi} e^{-\frac{1}{2}x^2}$, *called the* Gaussian function, *which plays an important role in statistics, it can be shown that its primitive,*

$$\int \frac{1}{2\pi} e^{-\frac{1}{2}x^2} \, dx,$$

cannot be expressed as a finite sum of elementary arithmetic expressions in terms of elementary functions $\exp, \ln, \sin, \sqrt{\ }$, *etc.*

Fortunately, in the *relatively* simple cases, the rules of differentiation help a lot in breaking down a function to recognizable derivatives.

The following are the most elementary rules for indefinite integration.

(10) $\qquad \displaystyle\int c \cdot f(x) \, dx = c \cdot \int f(x) \, dx \qquad\qquad\qquad$ constant multiple rule

(11) $\qquad \displaystyle\int [f(x) + g(x)] \, dx = \int f(x) \, dx + \int g(x) \, dx \qquad\qquad$ sum rule

These are direct consequences of the rules of differentation: If $F(x) = \int f(x)dx$, then $F'(x) = f(x)$ by definition, and by the constant multiple rule of differentiation, $[cF(x)]' = c[F(x)]' = cF'(x) = cf(x)$, and so $cF(x)$ is a primitive of $cf(x)$. If also $G(x) = \int g(x)dx$, and therefore $G'(x) = g(x)$, then by the sum rule of differentation $[F(x)+G(x)]' = F'(x)+G'(x) = f(x)+g(x)$. Consequently, $F(x) + G(x)$ is a primitive of the function $f(x) + g(x)$.

The power rule gives

$$\int x^r \, dx = \frac{1}{r+1} x^{r+1} + C, \qquad \text{if } r \neq -1,$$

because

$$\frac{d}{dx} \left(\frac{1}{r+1} x^{r+1} + C \right) = \frac{1}{r+1} \left(\frac{d}{dx} x^{r+1} \right) = \frac{1}{r+1} (r+1) x^r = x^r.$$

As a special case of this rule,

$$\int dx = \int 1 \, dx = \int x^0 \, dx = \frac{1}{1} x^{0+1} + C = x + C.$$

In the same way we can verify

$$(12) \qquad \int e^{kx} = \frac{1}{k} e^{kx} + C, \quad (k \neq 0)$$

$$(13) \qquad \int \frac{1}{x} \, dx = \ln |x| + C, \quad (x \neq 0)$$

$$(14) \qquad \int a^x \, dx = \frac{a^x}{\ln a} + C, \quad (a > 0)$$

$$(15) \qquad \int \sin x \, dx = - \cos x + C,$$

$$(16) \qquad \int \cos x \, dx = \sin x + C,$$

$$(17) \qquad \int \frac{1}{\cos(x)^2} \, dx = \tan x + C.$$

Exercise 17. *As an exercise in differentiation, verify equations* (12) *through* (17).

At table such as the above is called a *table of standard integrals* (or *standard primitives*). Because integration is so difficult, there are many large tables of standard integrals. Entire books are published that list known integrals, indexed according to some classification scheme of their integrand. Also, computer programs are available that have these tables built inn, and that will try to rewrite the integrand in such a form that it matches one of the internally tabulated standard integrals.[1]

The availability of computer programs does not dismiss us from gaining some skill in finding primitives ourselves—these computer programs have their limitations.

Example 46. *Find the primitive of $f(x) = \sqrt[3]{x}$.*

ANSWER. We need to find $\int \sqrt[3]{x} \, dx = \int x^{1/3} \, dx$ which, represented this way, is in the form of the standard integral $\int x^r \, dx$. Therefore, $F(x) = \int x^{1/3} \, dx = \frac{1}{1/3+1} x^{1/3+1} + C = \frac{3}{4} x^{4/3} + C$. To verify this answer, we differentiate: $F'(x) = \left[\frac{3}{4} x^{4/3} + C \right]' = \frac{3}{4} [x^{4/3}]' + 0 = \frac{3}{4} \frac{4}{3} x^{4/3-1} = x^{1/3} = f(x)$. $\qquad \square$

Example 47. *Find $\int \frac{1}{x^2} \, dx$.*

ANSWER. Rewrite $\int \frac{1}{x^2} \, dx = \int x^{-2} \, dx$ which is now in the standard form $\int x^r \, dx$, from which $F(x) = \int x^{-2} \, dx = \frac{1}{-2+1} x^{-2+1} + C = -\frac{1}{x} + C$. To verify, $F'(x) = [-1/x + C]' = -[1/x] + 0 = -(-1)x^{-2} = \frac{1}{x^2}$, which is the integrand. $\qquad \square$

Example 48. *Compute $\int (x^{-3} + 7e^{5x} + \frac{4}{x}) \, dx$.*

ANSWER. Use the sum and constant multiple rules to break the integral apart:

$$\int (x^{-3} + 7e^{5x} + \tfrac{4}{x}) \, dx = \int x^{-3} \, dx + 7 \int e^{5x} \, dx + 4 \int \tfrac{1}{x} \, dx.$$

---

[1] One such program, that is freely available for download on the internet is called <u>Maxima</u>

Each of the terms is now in a standard form, and so we read off the appropriate primitives:

$$\int (x^{-3} + 7e^{5x} + \tfrac{4}{x})\,dx = \frac{1}{-3+1}\,x^{-3+1} + 7\frac{1}{5}e^{5x} + 4\ln|x| + C$$

$$= -\frac{1}{2x^2} + \frac{7}{5}e^{5x} + 4\ln|x| + C,$$

where all constants of integration were absorbed into a single constant (as should always be done). To verify, $\left[-\frac{1}{2x^2} + \frac{7}{5}e^{5x} + 4\ln|x| + C\right]' = -\frac{1}{2}[x^{-2}]' + \frac{7}{5}[e^{5x}]' + 4[\ln|x|]' + 0 = -\frac{1}{2}(-2)x^{-3} + \frac{7}{5}5e^{5x} + 4\frac{1}{x} = x^{-3} + 7e^{5x} + \frac{4}{x}$, which is the integrand. $\qquad\square$

Note that because integration is an inferential process, the only way to actually make sure that the function found is a primitive, is through differentiation. We therefore *always* verify a solution by differentiating it. Failing to do so is effectively failing to show that the function provided is primitive.

2. Definite integration

As admitted earlier, the word "integration" is more related calculating areas than to anti-differentiation. As you can only calculate areas from clearly circumscribed definite regions, this process is called *definite integration*.

Suppose we have a continuous function $f(x)$, which is positive and bounded on the interval $[a, b]$. We would like to be able to compute the area under the graph of $f(x)$ as sketched in figure 1.
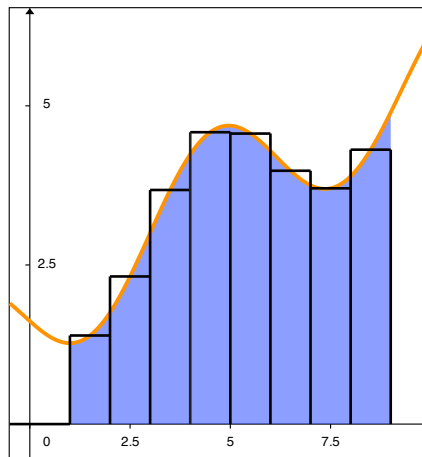

FIGURE 1. Area under the curve of $f(x)$: By this we mean "the area between the curve of the graph of $f(x)$ and the horizontal axis". The rectangles indicate a Rieman sum approximation.

We can do this approximately by dividing up the interval $[a, b]$ into n equal sized segments s_i, $i = 1 \ldots n$, of length $\Delta x = \frac{b-a}{n}$ each, and on each segment impose a rectangle as indicated in the figure. To determine the height of each rectangle, choose a point x_i (any point will do) in each segment s_i, $i = 1 \ldots n$. On the i-th segment the area under the curve is approximated by the area of a rectangle of width Δx and height $f(x_i)$, $f(x_i)\Delta x$. We now approximate the area on the

entire interval $[a, b]$ by simply summing the areas of the rectangles:

$$\text{area under curve} \approx f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{i=1}^{n} f(x_i)\,\Delta x.$$

This sum is called a *Riemann sum.* Clearly the Riemann sum becomes a more accurate approximation to the true area under the curve if we take more segments—i.e., if we make n larger, and hence, the width, $\Delta x = (b - a)/n$ smaller. In the limit, as $n \to \infty$, the approximation becomes arbitrarily accurate. We now define the symbol $\int_a^b f(x)dx$ to indicate this limit; that is,

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

by definition.

Consistent with geometric intuition, $\int_a^a f(x)dx = 0$ (the area underneath the curve always goes to 0 as the interval of integration becomes smaller and smaller). Furthermore, for $a < c < b$,

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

which is geometrically simply means that the area underneath the curve on the interval $[a, b)$ is equal to the sum of the area underneath the curve on the intervals $[a, c)$ and $[c, b)$.

Note that the only difference between the symbol $\int_a^b f(x)\,dx$ defined here as the limit of a Riemann sum, and the symbol $\int f(x)\,dx$ used in the previous section to denote the anti-derivative of $f(x)$, are the a and b attached to the $\int$ sign. These attachments are called the *integration limits* of the determined integral. The number a is the lower limit, the number b is the upper limit. In this case the dx term shows with respect to which variable the limit sum is calculated. In both cases $f(x)$ is called the integrand of the integral. Even though these two notations look almost identical, they signify a *very disparate* set of processes, although the Fundamental Theorem of Calculus ties these different concepts together.
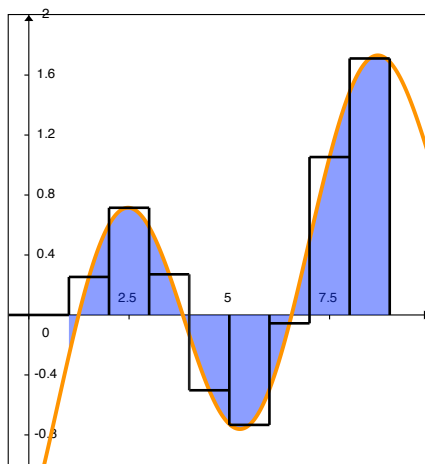

FIGURE 2. Area under the curve of $f(x)$ with negative parts. The rectangles indicate a Rieman sum approximation. The integral is defined with negative area.

Before turning to the Fundamental Theorem, first, let us consider functions that are not necessarily positive, such as the one sketched in figure 2. In this case we have to make a clear distinction between "area under the curve" (i.e., the physical area that is delimited by the graph of $f(x)$ and the horizontal axis in the range between the lower and upper integration limits), and

"the integral of $f(x)$ on the interval $[a, b]$": In the former case, since area is a positive quantity, the Riemann sums in the parts where $f(x)$ is negative should be calculated with $|f(x_i)|\Delta x$ as the rectangle area. The integral however is defined as earlier, and hence counts area that lies below the horizontal axis negatively! That is,

$$\text{the area between } a \text{ and } b = \lim_{n \to \infty} \sum_{i=1}^{n} |f(x_i)|\Delta x,$$

while

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

by definition. The integral thus conceptualizes the area blow the horizontal axis as *negative area*. This is on purpose, as it has many advantages. For instance, the concept of negative area allows us to define integrals $\int_a^b f(x)dx$ if $a > b$ as

$$\int_a^b f(x)dx = -\int_b^a f(x)dx,$$

which implies that the rule

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

holds for any three values $a, b, c \in \mathbb{R}$ (for which not necessarily $a < c < b$).
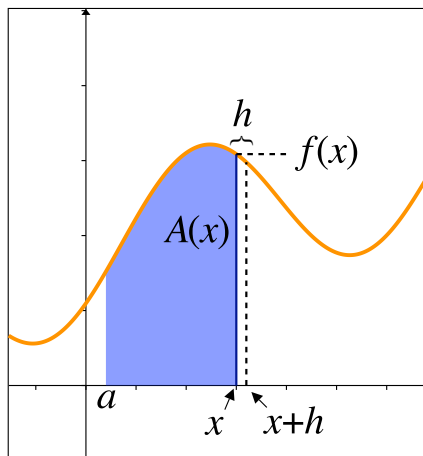

FIGURE 3. Intuitive explanation of the Fundamental Theorem of Calculus.

Fundamental Theorem of Calculus. We now turn to the intricate relation between anti-differentiation and integration. Let $f(x)$ be a function, and let $F(x)$ be a primitive of f; i.e., $F'(x) = f(x)$. The Fundamental Theorem (of Calculus) states that

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

This seemingly innocent result means that you can calculate areas under curves by finding primitives. We won't actually proof this theorem here, but to give some intuition.

Consider the shaded area under the curve in figure 3 between a and x. Suppose we have the function $A(x)$ which gives us the area of this region. Lets ask ourselves what the area under the curve should be between a and $x + h$? for some small h. That is, what is $A(x + h)$? We could approximate this area by the area of the shaded region, plus the area of a rectangle of height $f(x)$ and width h, that is,

$$A(x + h) \approx A(x) + f(x)h.$$

Off course, this approximation gets better and better as we take $h \to 0$. Now, lets compute the derivative of $A(x)$ in the point x. By definition of the derivative,

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

From the approximate equality this is approximately

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

In other words, the function $A(x)$ is a primitive of $f(x)$! Let $F(x)$ be another primitive of $f(x)$, then, due to the constant of integration of anti-derivatives, it must be the case that for some $C \in \mathbb{R}$,

$$A(x) = F(x) + C,$$

Hence, since $A(a) = 0$ (the shade area is then 0),

$$\int_a^x f(u)\, du = A(x) = A(x) - A(a) = F(x) + C - (F(a) - C) = F(x) - F(a),$$

which is precisely what is stated in the Fundamental Theorem of Calculus.

3. Applications

Solids of revolution.

Exercises

This section contains various exercises on anti-differentiation, using all the rules discussed so far.
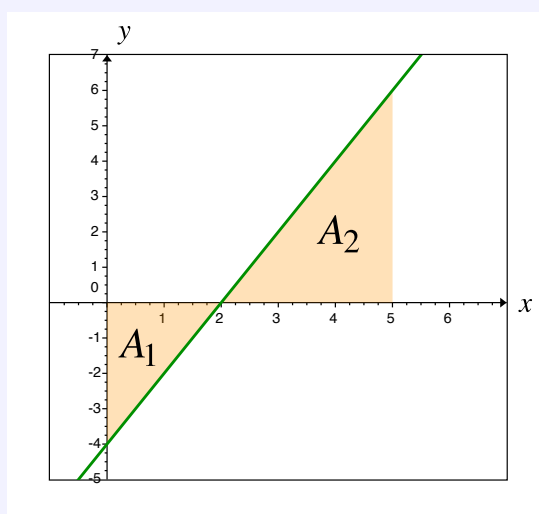
Example 49. *Compute $\int_0^5 (2x - 4)\, dx$.*

ANSWER.

$$\int_0^5 (2x - 4)\, dx = \left[x^2 - 4x\right]_0^5 = (5^2 - 4 \cdot 5) - (0^2 - 4 \cdot 0) = 25 - 20 = 5.$$

This can also be seen geometrically: The (signed) area of the shaded region is

$$A_2 - A_1 = \frac{1}{2} \cdot 3 \cdot 6 - \frac{1}{2} \cdots 2 \cdot 4 = 9 - 4 = 5.$$
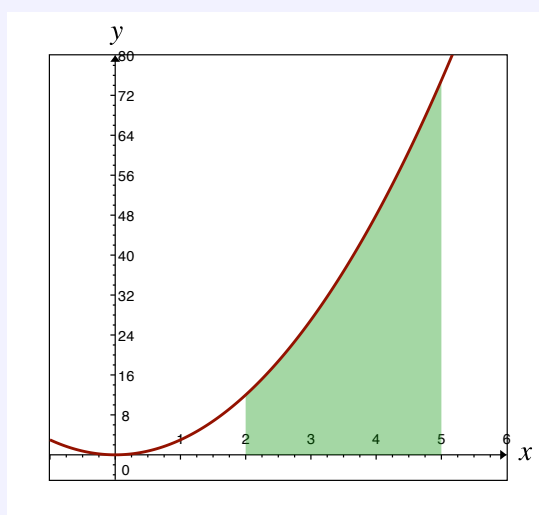
Example 50. *Compute $\int_2^5 3x^2\,dx$.*

ANSWER.

$$\int_2^5 3x^2\,dx = \left[x^3\right]_2^5 = 5^3 - 2^3 = 125 - 8 = 117.$$
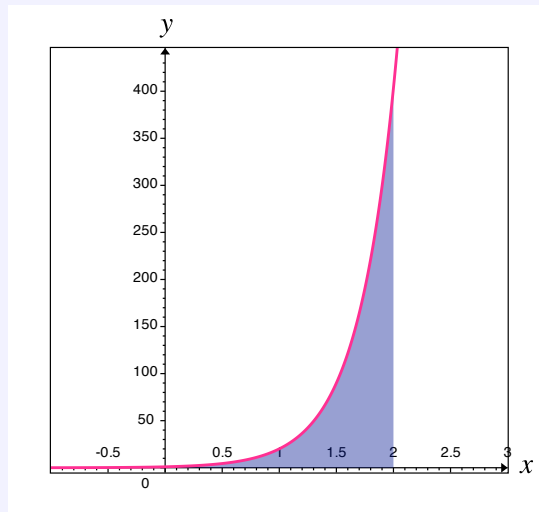
It is impossible to do this geometrically:

Example 51. *Compute $\int_0^2 e^{3x}\,dx$.*

ANSWER. $\int_0^2 e^{3x}\,dx = \left[\frac{1}{3}e^{3x}\right]_0^2 = \frac{1}{3}e^{3\cdot 2} - \frac{1}{3}e^{3\cdot 0} = \frac{1}{3}(e^6 - 1).$

Exercise 18. *Determine the following primitives*

(1) $\int 3^2 dx$

(2) $\int (x^2 - 3x + 2) dx$

(3) $\int \sqrt{x+1} dx$

(4) $\int \frac{1}{x+4} dx$

(5) $\int (2x+4)^7 dx$

(6) $\int \left(\frac{5}{x} - \frac{x}{5} \right) dx$

(7) $\int \frac{1}{\sqrt{x-7}} dx$

(8) $\int_{\ln 2}^{\ln 3} (e^x + e^{-x}) \, dx$

(9) $\int_0^{\ln 3} \frac{e^x + e^{-x}}{e^{2x}} dx$

Exercise 19. *Find the area under the curve $y = (3x - 2)^{-3}$ from $x = 1$ to $x = 2$.*

Chapter 4

Multivariable calculus

In this chapter we generalize the operations of differentiation and integration to functions of multiple variables. Functions of multiple variables map points in a multidimensional space onto points on a uni-dimensional space. That is, they have multiple inputs, and one output for each set of inputs. We will only consider functions that take numbers as both in and output. Mathematically the domain of a function of d variables, each of which takes values in $\mathbb{R}$, is denoted $\mathbb{R}^d$. The range of the function is usually $\mathbb{R}$. Often, the notations

$$f : \mathbb{R}^d \to \mathbb{R}, \qquad \text{and} \qquad (x_1, \ldots, x_d) \mapsto f(x_1, x_2, \ldots, x_d)$$

are used.

Example 52. *The function $f(x, y) = \frac{x^2 - y^2}{\frac{1}{10} + x^2 + y^2}$ is graphically displayed in figure 2. The domain of this function is $\mathbb{R}^2$ (the xy-plane or floor in the graph), and its range is $\mathbb{R}$ (the z-axis or height in the graph).*

Drawing graphs of functions of two variables is difficult when you can't draw well. One way to simplify this task is to draw *sections*.

Example 53 (Example 52 continued). *A y-section of the function $f(x, y)$ is the function that is obtained when the y variable is set to a fixed values, e.g.,*

$$g_0(x) = f(x, 0) = \frac{x^2}{\frac{1}{10} + x^2},$$

where y is fixed at $y = 0$. Other sections are $g_1(x) = f(x, 1)$, $g_3(x) = f(x, 3)$ etc. The following figure depicts several y sections of $f(x, y)$.

Graphs of functions of three variables, i.e., $(x, y, z) \mapsto f(x, y, z)$, are hard to imagine, and can only display *sections* of a four dimensional shape, by setting one of the variables fixed to a specific number, or, by plotting the contour (called an *isocline*) of the set of points for which the function has a particular value (e.g., for the function $f(x, y, z) = 3x^2 + 2y^2 + (y - 3)^2 z^2$ the isocline for $f(x, y, z) = 3$) is a spheroid like object—see figure 2). Similarly, functions of more than three variables can only be graphically depicted as sections in which all but two variables are set to fixed values, or by drawing contours keeping all but three variables fixed at given values.

In this chapter we will limit the discussion to functions of two variables. This is mainly to limit the notation. The techniques discussed in this chapter directly translate to functions of more than two variables.
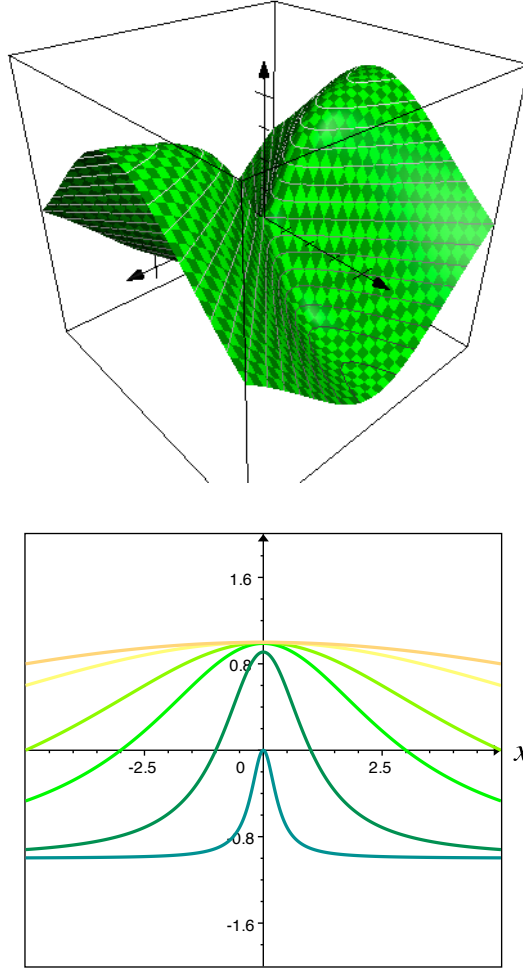
FIGURE 1. *Left*: Plot of the surface defined by $f(x, y)$ defined in Example 52 (left). *Right*: Graphs of *y-sections* of $f(x, y)$, for $y = 0$ (sea green), $1, 3, 5, 10, 15$ (orange).

1. Limits

Recall that to determine (the existence of) the limit $\lim_{x \to a} f(x)$ we had to make sure that it didn't matter whether x approached a from the left or from the right. In this case, therefore, x can approach a point a from only *two* directions. With a function of two variables, $f(x, y)$, the natural generalization of the limit of $f(x, y)$, as $(x, y) \to (a_x, a_y)$, denoted

$$\lim_{x, y \to a_x, a_y} f(x, y) = L,$$

maintains that L is only a limit, if $f(x, y)$ gets arbitrarily close to L as (x, y) approaches (a_x, a_y), *and* it doesn't matter from what direction (x, y) approaches (a_x, a_y). In this case, (x, y) can approach (a_x, a_y) from an *infinite* amount of directions. This is illustrated in figure 3, where the function $f(x, y) = (x^2 - y^2)/(10 + x^2 + y^2)$ are depicted by graphing its isoclines. If the relation $\lim_{x, y \to a_x, a_y} f(x, y) = f(a_x, a_y)$ holds, $f(x, y)$ is said to be continuous in the point (a_x, a_y). If $f(x, y)$ is continuous in all points of its domain, $f(x, y)$ is called *uniformly continuous*.
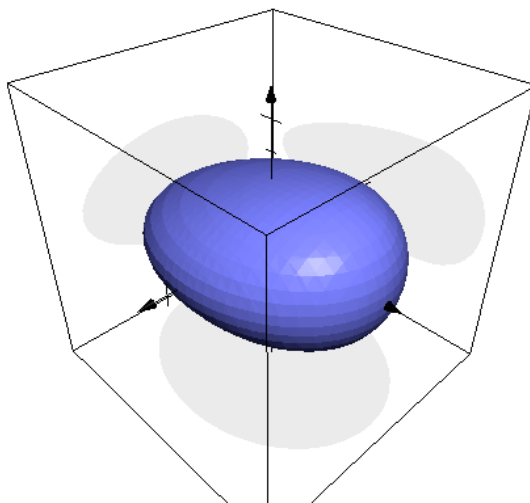
FIGURE 2. Contour (*isocline*) plot of function $f(x) = 3x^2 + 2y^2 + (y-2)^2 z^2$: the set of points (x, y, z) for which $f(x, y, z) = 5/2$ is displayed as a surface in 3D space.
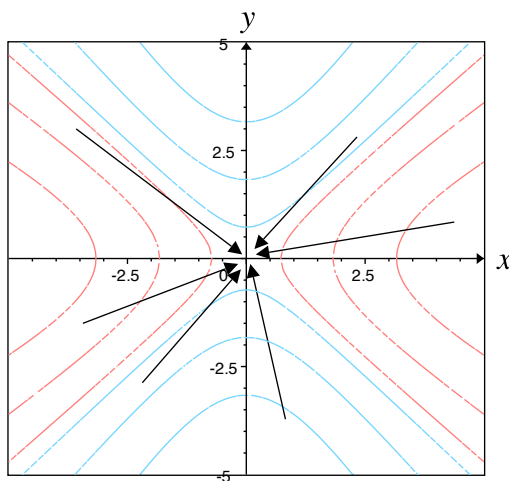

FIGURE 3. Isoclines graph for the function $f(x, y) = (x^2 - y^2)/(10 + x^2 + y^2)$, for $f(x, y) = \pm 1/2, \pm 1/4, \pm 1/20$ (dashed lines). For $f(x, y) > 0$ the isoclines are colored red. For $f(x, y) < 0$ the isoclines are blue. (In 3D, this function looks a lot like the one in the left panel of figure 2.) The limit of $f(x, y)$ where (x, y) approach the point $(0, 0)$ only exists if its value is independent of the direction from which $a(0, 0)$ is approached.

2. Derivatives

The discussion on limits of multivariable functions reveal that the direction of approach of a point matters. Unlike uni-variable functions, there's no obvious unique definition if our concept of derivative for these functions is to be consistent with the idea of steepness of 'the tangent' to 'the curve' of the graph; a glance at the right hand panel of figure 2 shows that the steepness of the surface depends on the direction in which one is traveling. This is no different from walking on the slope of a mountain: the steepness of the path depends on the direction one is heading.

This problem is resolved by considering the idea of sections again. Let $f(x, y)$ be a function of two variables, and let

$$u_y(x) = f(x, y)$$

be a y-section of $f(x, y)$. Then the *partial derivative* of $f(x, y)$ with respect to x is defined as

$$f_x(x, y) = \frac{\partial f}{\partial x}(x, y) = \frac{d}{dx} u_y(x) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}.$$

Similarly, the partial derivative of $f(x, y)$ with respect to y is defined as

$$f_y(x, y) = \frac{\partial f}{\partial y}(x, y) = \frac{d}{dy} v_x(y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h},$$

where $v_x(y) = f(x, y)$ is an x-section of $f(x, y)$. The symbol ∂ is used to distinguish it from the *total derivative* that is to be discussed shortly.

Example 54. *Let*
$$f(x, y) = x^2 y^3 + 3x + 2y + 6.$$
Find the partial derivatives of $f(x, y)$ with respect to x and y.

ANSWER. The partial derivatives are given by differentiating $f(x, y)$ with respect to x or y while considering the other variable as a constant: For x,

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{d}{dx} \left(x^2 y^3 + 3x + 2y + 6 \right) \\
&= y^3 \frac{d}{dx} (x^2) + 3 \frac{d}{dx} (x) + \frac{d}{dx} (2y) + \frac{d}{dx} (6) \\
&= y^3 (2x) + 3 + 0 + 0 \\
&= 2xy^3 + 3.
\end{aligned}$$

For y,

$$\begin{aligned}
\frac{\partial f}{\partial y} &= \frac{d}{dy} \left(x^2 y^3 + 3x + 2y + 6 \right) \\
&= x^2 \frac{d}{dy} (y^3) + \frac{d}{dy} (3x) + \frac{d}{dy} (2y) + 6 \\
&= x^2 (3y^2) + 0 + 2 + 0 \\
&= 3x^2 y^2 + 2.
\end{aligned}$$

Hence, the partials are $f_x(x, y) = 2xy^3$, and $f_y(x, y) = 3x^2 y^2 + 2$.   □

The partial derivative is analogously defined for functions of more than 2 variables, by considering all but one variable as fixed constants.

How are partial derivatives related to steepness? It turns out that if we consider the point $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ in the xy-plain, and draw an arrow from the origin (the point $(0, 0)$) to this point, then the arrow is pointing into the direction of steepest ascend. In the analogy of the mountain slope, if we measure the inclination in the x direction (that is, $\partial f / dx$) and in the inclination in the y-direction, and draw an arrow from the point where we are standing to $(\partial f / \partial x, \partial f / \partial y)$, then that arrow points in the direction opposite to which a ball would role, or water would flow. We won't prove this point here; we do remark however that the vector $(\partial f / \partial x, \partial f / \partial y)$ is called the *gradient* of $f(x, y)$ and is often denoted with the special symbol

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}.$$

The same observation and notation extends naturally to functions of more than two variables. (Only the analogy with the mountain slope becomes a bit harder to imagine.)

In the same way as it was the case with normal derivatives, partial derivatives do not necessarily exist at a given point. A necessary (but insufficient) condition for the derivative to exist is that the function is continuous in the point under consideration.

Sometimes the partial derivatives themselves can be differentiated with respect to all the variables.

Example 55 (Example 54 continued). *Consider $f(x,y)$ of Example 54. Find the second order partial derivatives.*

ANSWER. The partial derivative of $f(x,y)$ with respect to x, $f_x(x,y)$ can be differentiated with respect to x again, to obtain

$$f_{xx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(2xy^3 + 3\right) = 2y^3.$$

The partial derivative $f_x(x,y)$ can also be differentiated with respect to y,

$$f_{xy}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}\left(2xy^3 + 3\right) = 6xy^2.$$

Similarly, we can differentiate $f_y(x,y)$;

$$f_{yy}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(3x^2 y^2 + 2\right) = 6x^2 y,$$

and

$$f_{yx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial x}\left(3x^2 y^2 + 2\right) = 6xy^2.$$

□

In the last example we found that

$$f_{xy}(x,y) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = f_{yx}(x,y).$$

This is by no means an accident. In general, if $f_{xy}(x,y)$ and $f_{yx}(x,y)$ are both continuous, it is true that $f_{xy}(x,y) = f_{yx}(x,y)$.

Chain rule. Without proof we state the chain rule for functions of two variables. The generalization to more that two variables is, again, straightforward. Give functions $f(x,y)$, $g(t)$ and $h(t)$, the chain rule for the composite function $f\left((g(t), h(t)\right)$ is

$$\frac{d}{dt} f\left(g(t), h(t)\right) = \frac{\partial f}{\partial x}\left(g(t), h(t)\right) \frac{d}{dt} g(t) + \frac{\partial f}{\partial y}\left(g(t), h(t)\right) \frac{d}{dt} h(t).$$

Example 56. *Let $g(t) = \frac{(t^3 + 4)^2}{(t-2)(t+2)}$. Find $g'(t)$.*

ANSWER. Define $f(x, y) = x/y$, $x(t) = (t^3 + 4)^2$, and $y(t) = (t - 2)(t + 2) = t^2 - 4$. Then,

$$\frac{\partial f}{\partial x} = \frac{1}{y}, \quad \text{and} \quad \frac{\partial f}{\partial y} = -\frac{x}{y^2},$$

and

$$x'(t) = 6t^2(t^3 + 4), \quad \text{and} \quad y'(t) = 2t.$$

Hence, with the chain rule for multiple variable functions,

$$g'(t) = \frac{1}{y} 6t^2(t^3 + 4) - \frac{x}{y^2} 2t = \frac{6t^2(t^3 + 4)}{(t - 2)(t + 2)} - \frac{2t(t^3 + 4)^2}{(t - 2)^2(t + 2)^2}.$$

The same result is obtained through direct differentiation of $g(t)$.   □

Example 57. *As a special case, if $f(x, y)$ is a function of x and y, and in addition y is a function of x. Then*

$$\frac{d}{dx} f(x, y(x)) = \frac{\partial f}{\partial x}(x, y(x)) \frac{d}{dx}(x) + \frac{\partial f}{\partial y}(x, y(x)) \frac{d}{dx} y(x)$$

$$= \frac{\partial f}{\partial x}(x, y(x)) + \frac{\partial f}{\partial y}(x, y(x)) \frac{d}{dx} y(x),$$

which is called the total derivative *of $f(x, y)$ with respect to x, as opposed to the partial derivative $\frac{\partial f}{\partial x}$.*

3. Exercises

Exercise 20. *Determine all the first-order partial derivatives of*

(1) $f(x, y) = 2x^3 y + 3xy + \dfrac{y}{x}$

(2) $f(x, y) = (xy^2 + 1)^5$

(3) $f(x, y) = xye^{xy}$

(4) $f(x, y) = \ln\left(\dfrac{xy}{x + 3y}\right)$

Exercise 21. *Compute f_{xx}, f_{xy}, f_{yx} and f_{yy} of*

(1) $f(x, y) = x^2 + y^3 - 2xy^2$

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

(3) $f(x, y) = x \ln y$

4. Extreme values & Optimization

Recall that the derivative of a function can be used to find extrema—maxima and minima—of functions: At the maximum or minimum of a function the slope of the tangents is zero, while just left of a maximum (minimum) the slope is positive (negative) which just right of the maximum (minimum) the slope is negative (positive). The latter means that in a (small) neighborhood around the maximum (minimum) the slopes are decreasing (increasing) as we move from left to

right, and so the slope is an decreasing (increasing) function in the neighborhood of the maximum (minimum), implying that the second order derivative must be negative (positive) in that neighborhood.

Following directly from these observations, we found the recipe for finding maxima (minima) of a function $f(x)$:

(1) Compute the derivative $f'(x)$.
(2) Solve the equation $f'(x) = 0$ for x.
(3) Compute the second order derivative $f''(x)$
(4) Evaluate $f''(x)$ in the point found in step 2
   - If $f''(x) > 0$, then a minimum is found.
   - If $f''(x) < 0$, then a maximum is found.
   - If $f''(x) = 0$, then the point found may be a minimum, a maxim, or neither (a *saddle point*).

This recipe generalizes to the multivariable case as follows: Recall that the gradient of $f(x, y)$,

$$\nabla f(x, y) = \left(\begin{array}{c} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{array} \right),$$

considered as an vector (arrow), is pointing in the direction of steepest ascend—that is, opposite to the direction that a ball on the slope of a mountain would role. Then it is clear that if it is zero, i.e., $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$, then in no direction is it possible to ascend or descend (at least not for an infinitesimal step size)—the ball will not start to role from that spot on the mountain. This can only occur in a maximum, a minimum or a saddle point (illustrated in figure 4). It can also be reasoned from the sections of the function that at least it must be that $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$ at a minimum or maximum, otherwise the tangents to those sections would have a slope unequal to zero.
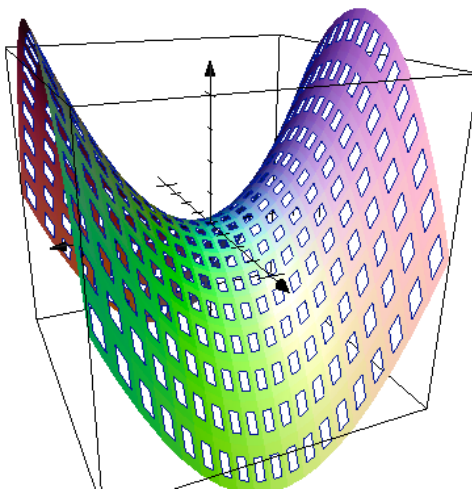
FIGURE 4. Example of a saddle point. The name obviously derived from the shape.

From the saddle point case, it is clear that $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$ are not sufficient conditions. The second condition that must be satisfied concerns the matrix of second order partial derivatives: For a function of two variables, $f(x, y)$, let

$$\Delta(x, y) = \det \left[\begin{array}{cc} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{array} \right] = f_{xx}(x, y) f_{yy}(x, y) - f_{xy}(x, y)^2$$

be the determinant of the Hessian matrix of $f(x, y)$. Let (x_0, y_0) solve the equations

$$\frac{\partial f}{\partial x}(x, y) = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(x, y) = 0.$$

Then,

- If $\Delta(x_0, y_0) > 0$ *and* $f_{xx}(x_0, y_0) < 0$, then $f(x, y)$ has a local maximum at (x_0, y_0).
- If $\Delta(x_0, y_0) > 0$ *and* $f_{xx}(x_0, y_0) > 0$, then $f(x, y)$ has a local minimum at (x_0, y_0).
- If $\Delta(x_0, y_0) < 0$, then $f(x, y)$ has a saddle point at (x_0, y_0).
- If $\Delta(x_0, y_0) = 0$, then $f(x, y)$ may have maximum, minimum, or a saddle point at (x_0, y_0).

Example 58. *Let* $f(x, y) = \dfrac{1}{2}x^2 + y^2 - 3x + 2y - 8$. *Find the critical points of* $f(x, y)$ *(i.e., find maxima, minima, and/or, saddle points).*

ANSWER. First we solve the equations

$$f_x(x, y) = \frac{\partial}{\partial x}\left(\frac{1}{2}x^2 + y^2 - 3x + 2y - 8\right) = x - 3 = 0 \qquad \text{and}$$

$$f_y(x, y) = \frac{\partial}{\partial y}\left(x^2 + y^2 - 3x + 2y - 8\right) = 2y + 2 = 0,$$

to find $x_0 = 3$, and $y_0 = -1$. So, $f(x, y)$ has one critical point at $(3, -1)$. To see what kind of critical point, we check the determinant of the Hessian matrix. The second order partials are,

$$f_{xx}(x, y) = \frac{\partial}{\partial x}(x - 3) = 1 \quad f_{xy}(x, y) = \frac{\partial}{\partial x}(2y + 2) = 0$$
$$f_{yx}(x, y) = f_{x,y}(x, y) = 0 \quad f_{yy}(x, y) = \frac{\partial}{\partial y}(2y + 2) = 2 \,.$$

Therefore, the determinant is

$$\Delta(x, y) = f_{xx}(x, y)f_{yy}(x, y) - f_{x,y}(x, y)^2 = 1 \cdot 2 = 2.$$

Hence, $\Delta(x_0, y_0) > 2$, and so, since $f_{xx}(x_0, y_0) = 1 > 0$, $f(x, y)$ has a minimum at $(3, -1)$.   □

Example 59 (Least Squares). *Suppose we have obtained pairs* (x_i, y_i), $i = 1, \ldots, n$ *in an empirical study. A scatter plot of these pairs is given below. In ordinary linear regression, a straight line is fitted to the points such that the squared differences between the points and their prediction from the line is minimized.*

SOLUTION. More formally: For each point (x_i, y_i) we have the model

$$y_i = \alpha + \beta\, x_i + \epsilon_i,$$

where ϵ_i is the prediction error, and we want to minimize the function (of the two variables α and β)

$$S = \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2 = \sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}(y_i - \alpha - \beta\, x_i)^2.$$

We need to minimize $S(\alpha, \beta)$ with respect to α and β *(why?)*. To do so, we first set the partial derivatives equal to zero and solve for α and β, and then check, using the second order partial derivatives, whether a minimum was obtained.

The partial derivatives are

$$\frac{\partial S}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^n (y_i - \alpha - \beta\, x_i)^2 = \sum_{i=1}^n \frac{\partial}{\partial \alpha} (y_i - \alpha - \beta\, x_i)^2 = \sum_{i=1}^n 2(y_i - \alpha - \beta\, x_i)(-1),$$

$$\frac{\partial S}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^n (y_i - \alpha - \beta\, x_i)^2 = \sum_{i=1}^n \frac{\partial}{\partial \beta} (y_i - \alpha - \beta\, x_i)^2 = \sum_{i=1}^n 2(y_i - \alpha - \beta\, x_i)(-x_i).$$

(*Note the exchange of the $\sum$ and $\frac{\partial}{\partial x}$ symbols—is it valid?*) Setting these equal to zero, gives the equations (*check it!*)

$$-2\sum_{i=1}^n (y_i - \alpha - \beta\, x_i) = 0 \iff \alpha n + \beta \left(\sum x_i\right) = \left(\sum y_i\right)$$

$$-2\sum_{i=1}^n (y_i x_i - \alpha x_i - \beta\, x_i^2) = 0 \iff \alpha \left(\sum x_i\right) + \beta \left(\sum x_i^2\right) = \left(\sum x_i y_i\right).$$

Solving for α and β gives

$$\alpha = \tfrac{1}{n} \left(\sum y_i\right) - \beta \tfrac{1}{n} \left(\sum x_i\right), \qquad \beta = \frac{\sum x_i y_i - \frac{1}{n} \left(\sum y_i\right) \left(\sum x_i\right)}{\sum x_i^2 - \frac{1}{n} \left(\sum x_i\right)^2}.$$

Next we check the determinant of the Hessian matrix, even though we have found a single critical point and since S is a sum of squares this critical point can only be a minimum. The second order partials are (*check it!*)

$$S_{\alpha\alpha} = \frac{\partial^2 S}{\partial \alpha^2} = 2n, \qquad S_{\alpha\beta} = \frac{\partial^2 S}{\partial \alpha \partial \beta} = 2\sum x_i, \qquad S_{\beta\beta} = \frac{\partial^2 S}{\partial \beta^2} = 2\sum x_i^2,$$

and so, the determinant of the Hessian matrix is

$$\Delta(\alpha, \beta) = S_{\alpha\alpha} S_{\beta\beta} - S_{\alpha\beta}^2 = 4\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right],$$

which is independent of α and β. It will be left as an exercise to show that $\Delta(\alpha, \beta) > 0$. $\qquad \square$

Exercise 22. *Show that $\Delta(\alpha, \beta) > 0$. Hint: Show that $n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 = n \sum_{i=1}^n (x_i - \overline{x})^2$, where $\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$.*

For functions of more than to variables, the situation is a bit different. Suppose we have $f(x_1, \ldots, x_d)$. Suppose $(x_{01}, x_{02}, \ldots, x_{0d})$ solve the equations

$$\frac{\partial f}{\partial x_1} = 0, \frac{\partial f}{\partial x_2} = 0, \ldots, \frac{\partial f}{\partial x_d} = 0.$$

Let

$$H(x_1, \ldots, x_d) = \begin{bmatrix} f_{x_1 x_1} & f_{x_1 x_2} & \cdots & f_{x_1 x_d} \\ f_{x_2 x_1} & f_{x_2 x_2} & & f_{x_2 x_d} \\ \vdots & & \ddots & \vdots \\ f_{x_d x_1} & \cdots & \cdots & f_{x_d x_d} \end{bmatrix}$$

be the Hessian matrix of $f(x_1, \ldots, x_d)$—that is, the matrix with second order partials of $f(x_1, \ldots, x_d)$. Then,

- If the eigenvalues of $H(x_{01}, \ldots, x_{0d})$ are all negative, then $f(x_1, \ldots, x_d)$ has a maximum at $(x_{01}, \ldots, x_{0d})$.
- If the eigenvalues of $H(x_{01}, \ldots, x_{0d})$ are all positive, then $f(x_1, \ldots, x_d)$ has a minimum at $(x_{01}, \ldots, x_{0d})$.
- If some of the eigenvalues of $H(x_{01}, \ldots, x_{0d})$ are positive and the rest are all negative, then $f(x_1, \ldots, x_d)$ has a saddle point at $(x_{01}, \ldots, x_{0d})$.
- If any of the eigenvalues of $H(x_{01}, \ldots, x_{0d})$ zero, then $f(x_1, \ldots, x_d)$ may have a maximum, minimum, or a saddle point in $(x_{01}, \ldots, x_{0d})$.

Example 60 (Heat loss). *Suppose that we have to design a rectangular building with a volume of 147,840 cubic feet. Assuming that the daily loss of heat is given by*

$$w = 11xy + 14yz + 15xz,$$

where x, y, and z are, respectively, the length, width, and height of the building, find the dimensions of the building for which the daily heat loss is minimal.

ANSWER. We must minimize the function

$$w = 11xy + 14yz + 15xz,$$

subject to the constraint that

$$V = xyz = 147,800 \text{ feet}.$$

The constraint implies that $z = V/(xy)$, which we can substitute in the objective function to obtain a minimization problem without constraint of a function of two variables, viz.,

$$g(x,y) = 11xy + 14yV/(xy) + 15xV/(xy) = 11xy + 14\frac{V}{x} + 15\frac{V}{y}.$$

To find the minima of this function, we equate the partial derivatives with zero,

$$g_x(x,y) = \frac{\partial}{\partial x}\left(11xy + 14\frac{V}{x} + 15\frac{V}{y}\right) = 11y - 14\frac{V}{x^2} = 0 \iff y = \frac{14}{11}\frac{V}{x^2},$$

$$g_y(x,y) = \frac{\partial}{\partial y}\left(11xy + 14\frac{V}{x} + 15\frac{V}{y}\right) = 11x - 15\frac{V}{y^2} = 0 \iff y = \pm\sqrt{\frac{15}{11}\frac{V}{x}}.$$

Clearly, y should be positive, and hence, we find an equation for x,

$$\frac{14}{11}\frac{V}{x^2} = \sqrt{\frac{15}{11}\frac{V}{x}},$$

which is solved for x by $x = (14^2 V/11 \cdot 15)^{1/3} = (14^2 \cdot 147,840/11 \cdot 15)^{1/3} = 56$. For y we therefore have $y = \frac{14}{11}\frac{V}{56^2} = 60$, and for z we find $z = V/(xy) = V/(60 \cdot 56) = 44$.

To verify that this is indeed a minimum, we compute the determinant of the Hessian matrix. The second order partials are

$$g_{xx}(x,y) = 28\frac{V}{x^3}, \quad g_{xy}(x,y) = g_{yx}(x,y) = 11, \quad \text{and} \quad g_{yy}(x,y) = 30\frac{V}{y^3},$$

and so, the determinant is

$$\Delta(x,y) = g_{xx}(x,y)g_{yy}(x,y) - g_{xy}(x,y)^2 = 28 \cdot 30 \cdot \frac{V^2}{x^3 y^3} - 121.$$

Evaluated in the critical point the determinant equals $\Delta(56,60) = 363 > 0$, and because this is positive and $g_{xx}(56,60) = 28V/56^3$ is positive, the point $(56,60)$ is indeed a minimum of $g(x,y)$.

In conclusion, we should build our building with dimensions $56 \times 60 \times 44$ feet, to minimize the heat loss.

5. Integration of functions of multiple variables

In this section we discuss multiple integration. That is, we discuss the situation in which we integrate over more than one variable. We confine the discussion to double integration, as it extends readily to higher order integration.

One motivation for integration was the determination of area under a curve. The obvious extension to functions of two variables is to calculate the volume underneath a surface. In the one variable case we determined the area through a process of evermore accurate approximation with rectangles whose heights were determined by the curve, and whose widths were determined by the number of rectangles and the integration limits (viz., the width of the integration interval).
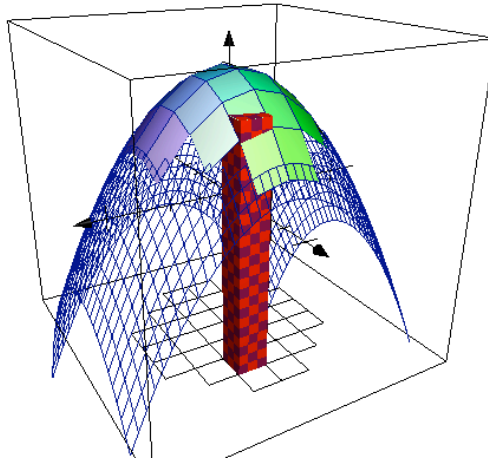
FIGURE 5. Approximation of volume underneath the surface specified by a function by rectangular blocks. The (disc shaped) region of integration is indicated by the grid in the xy-plane.

This process of increasingly accurate approximations readily translates in a similar process for functions of two variables. We sketch this process. Given a region R in the xy-plane over which the volume is to be determined, divide up the x-range of the region into n segments of equal length, and divide up the y-range of the region into m segments of equal length. Let these segments specify a grid of rectangles of area $\Delta x \Delta y$, where $\Delta x \to 0$ as $n \to \infty$, and $\Delta y \to 0$ as $m \to \infty$. Impose on each rectangle a block of dimensions $\Delta x \times \Delta y \times f(x_i, y_j)$, where (x_i, y_j) is an arbitrary point inside the i, j-th rectangle. Figure 5 illustrates this. Approximate the volume underneath the surface specified through $f(x, y)$ by the Riemann sum

$$V \approx \sum_{i=1}^{n} \sum_{j=0}^{m} f(x_i, y_j) 1_R(i, j) \Delta x \Delta y,$$

where the function $1_R(i, j)$ equals 1 if the i, j-th rectangle overlaps with the region of integration R and is zero otherwise. The double integral is now defined to be

$$\iint_R f(x, y) dx dy = \lim_{n,m \to \infty} \sum_{i=1}^{n} \sum_{j=0}^{m} f(x_i, y_j) 1_R(i, j) \Delta x \Delta y.$$

Suppose the region of integration can be specified as follows

$$R = \{(x, y) | a \leq x \leq b, g(x) \leq y \leq h(x)\},$$

which says that R consists of those points (x, y) in the xy-plane such that the x-coordinate lies between a and b, and for each x-coordinate, the corresponding y-coordinate lies between two know functions $g(x)$ and $h(x)$, $g(x) \leq h(x)$. A fundamental result (*Fubini's theorem*) now says that

$$\iint_R f(x, y) dx dy = \int_a^b \left(\int_{g(x)}^{h(x)} f(x, y) dy \right) dx.$$

That is, we can calculate the integral—determine the volume underneath de surface specified by $f(x, y)$ in the region R—by first integrating $f(x, y)$ with respect to y between $g(x)$ and $h(x)$, and then integrating the resulting function of x over the interval (a, b). Note that the result of the first, innermost, integral is indeed a function of x, since if $F(x, y)$ is a function such that

$\frac{d}{dy} F(x,y) = f(x,y)$, then, by the Fundamental Theorem of Calculus,

$$\int_{g(x)}^{h(x)} f(x,y)dy = F(x,h(x)) - F(x,g(x)),$$

which is clearly a function of x (and *only* x). Consequently,

$$\iint_R f(x,y)dxdy = \int_a^b F(x,h(x))dx - \int_a^b F(x,g(x))dx.$$

An expression of the form $\int_a^b \left(\int_{g(x)}^{h(x)} f(x,y)dy \right) dx$ is called an *iterated integral*.

Example 61. *Let R be the region $R = \{(x,y)|1 \leq x \leq 2, 3 \leq y \leq 4\}$. Compute*

$$\iint_R (y-x)dxdy$$

ANSWER. The double integral can be calculated as the iterated integral

$$\iint_R (y-x)dxdy = \int_1^2 \left(\int_3^4 (y-x)dy \right) dx.$$

We first compute the innermost integral.

$$\int_3^4 (y-x)dy = \int_3^4 ydy - \int_3^4 xdy = \left[\frac{1}{2}y^2\right]_3^4 - [xy]_3^4 = \frac{1}{2}(4^2 - 3^2) - (4x - 3x) = \frac{7}{2} - x,$$

which is a function of x. Next we compute the outermost integral using the function found,

$$\int_1^2 \left(\frac{7}{2} - x \right) dx = \left[\frac{7}{2}x\right]_1^2 - \left[\frac{1}{2}x^2\right]_1^2 = (7 - \frac{7}{2}) - (2 - 1/2) = 2.$$

Therefore, $\iint_R (y-x)dxdy = 2$.   □

Example 62. *Evaluate the iterated integral*

$$\int_0^1 \left(\int_{\sqrt{x}}^{x+1} 2xydy \right) dx.$$

SOLUTION. We first calculate the innermost integral.

$$\int_{\sqrt{x}}^{x+1} 2xydy = [xy^2]_{\sqrt{x}}^{x+1} = x((x+1)^2 - x) = x^3 + x^2 + x.$$

Which is a function of x, as to be expected. Next we compute the outermost integral,

$$\int_0^1 (x^3 + x^2 + x)dx = \left[\frac{1}{4}x^4 + \frac{1}{3}x^3 + \frac{1}{2}x^2\right]_0^1 = \frac{1}{4} + \frac{1}{3} + \frac{1}{2} = \frac{13}{12}.$$

□

Example 63. *Let R be given as*

$$R = \{(x,y)|0 \leq x \leq 2, 0 \leq y \leq x/2\}.$$

Compute

$$\iint_R e^{2y-x}dxdy.$$

ANSWER. The integral can be computed from the iterated integral

$$\iint_R e^{2y-x} dx dy = \int_0^2 \left(\int_0^{x/2} e^{2y-x} dy \right) dx.$$

We compute the innermost integral first,

$$\int_0^{x/2} e^{2y-x} dy = e^{-x} \int_0^{x/2} e^{2y} dy = e^{-x} \left[\frac{1}{2} e^{2y} \right]_0^{x/2} = e^{-x} (\frac{1}{2} e^x - \frac{1}{2}) = \frac{1}{2} - \frac{1}{2} e^{-x}.$$

This is a function of x, as it should be, and we substitute it into the outermost integral,

$$\int_0^2 \frac{1}{2}(1 - e^{-x}) dx = \frac{1}{2} \left[x + e^{-x} \right]_0^2 = \frac{1}{2}(2 + e^{-2} - 0 - 1) = \frac{1}{2} + \frac{1}{2} e^{-2}.$$

Hence, $\iint_R e^{2y-x} dx dy = \frac{1}{2} + \frac{1}{2} e^{-2}$ is the volume underneath e^{2y-x} in the indicated region. $\square$

6. Exercises

Exercise 23. *Let $R = \{(x,y) | 0 \le x \le 2, 2 \le y \le 3\}$, calculate the following double integrals.*

Interpret each as a volume.

(1) $\iint_R xy^2 dx dy$

(2) $\iint_R (xy + y^2) dx dy$

(3) $\iint_R e^{-x-y} dx dy$

(4) $\int_R e^{y-x} dx dy$

Exercise 24. *Let $f(x,y) = x^2 + y^2$. Calculate the volume bounded above by $f(x,y)$ over the*

region R, if R

(1) *is the rectangle bounded by the lines $x = 1$, $x = 3$, $y = 0$, $y = 3$.*

(2) *is the region $\{(x,y) | 0 \le x \le 1, 0 \le y \le \sqrt{x}\}$.*

Chapter 5

Integration Techniques

In this chapter we will discuss more systematic methods for finding primitives, than mere recognition. These are especially useful for more complex (looking) integrals.

Not all integrals allow for an explicit expression of the primitive of a function in terms of a finite number of basic arithmetic operations and transcendental functions and exponents (so called 'closed form' expressions). Some such integrals play an important role in many fields, such as statistics. Fortunately, any definite integral can be computed numerically, as we shall see.

1. Integration by substitution

The chain rule of differentiation reads

$$[g(h(x))]' = g'(h(x))h'(x).$$

Hence, $g(h(x))$ is a primitive of $g'(h(x))h'(x)$, and can, by definition of the indefinite integral, be written as

$$(18) \qquad g(h(x)) = \int g'(h(x))h'(x)dx.$$

If we, therefore, have an integral

$$\int f(x)dx,$$

where we recognize that $f(x)$ is of the form

$$f(x) = g'(h(x))\, h'(x),$$

then to find the primitive of $f(x)$, we really only need to find $g(y)$ on the left hand side of Equation (18)—the primitive of $g'(x)$—as we have already recognized $h(x)$ and $h'(x)$. That is, we really only need to find

$$g(y) = \int g'(y)dy,$$

and then substitute y for $h(x)$, yielding $g(h(x))$. Or,

$$\int g'(h(x))h'(x)dx = \int g(y)dy\Big|_{y=h(x)} \qquad \text{(substitution rule)},$$

which can be used to tremendously simplify integrals. Slightly more suggestive notation for this important rule is

$$\int f(y(x))\frac{dy}{dx}dx = \int f(y)dy, \quad \text{or} \quad \int g'(h(x))h'(x)dx = \int g'(h(x))dh(x).$$

Example 64. *Compute* $\int \dfrac{2x}{x^2+1}dx$.

ANSWER. We recognize the numerator, $2x$, to be the derivative of the denominator, $x^2 + 1$. Hence, if we rewrite into the form of the substitution rule

$$\int \frac{2x}{x^2+1}dx = \int \frac{1}{x^2+1}(x^2+1)'dx = \int \frac{1}{y}dy\Big|_{y=x^2+1}.$$

The last integral is the standard integral for the logarithm, and hence

$$\int \frac{2x}{x^2+1}dx = \left[\ln|y| + C\right]_{y=x^2+1} = \ln|x^2+1| + C.$$

As always, we check the result by differentiation, $[\ln|x^2+1| + C]' = \frac{1}{x^2+1}[x^2+1]' + 0 = \frac{2x}{x^2+1}$, which implies that our solution is indeed requested primitive. □

A key requirement of this technique is our ability to recognize $h(x)$ and $h'(x)$. This means, again, lots of practice with both differentiation as well as with integration. However, we can use the following trial and error technique that is also more easy to memorize than the substitution rule.

Example 65 (Example 64 revisited). *Compute* $\int \dfrac{2x}{x^2+1}dx$.

ANSWER. We *try and choose* u to substitute the numerator, i.e., $u = x^2 + 1$. First compute $\frac{du}{dx} = [x^2+1]' = 2x$, which we now *suggestively* write as

$$du = 2xdx.$$

We now substitute u for every occurence of the expression $x^2 + 1$, and du for the expression $2xdx$ in the integral above,

$$\int \frac{2x}{x^2+1}dx = \int \frac{1}{x^2+1}2xdx = \int \frac{1}{u}du = \log|u| + C.$$

We now need to substitute back $u = x^2 + 1$ into the resulting primitive, so that we find

$$\int \frac{2x}{x^2+1}dx = \ln|x^2+1| + C,$$

which is the same as before. □

As a general note of advice, it is (almost) always a good idea to let your choice of u capture as large a chunk as possible of the integrand expression. For instance, in the example above, we could have also tried $u = x^2$ first (which is a smaller part of the integrand than $u = x^2 + 1$), but then we would have ended up with the integral $\int \frac{1}{u+1}du$, which then needs to be integrated by substitution again. (In this case it is not too difficult to integrate by direct recognition, but this is not always the case.) By letting u capture as much as possible, integration by substitution can save you a lot work. If your choice of u doesn't work out the first time, you can try and choose a smaller chunck.

Example 66. *Compute* $\int \frac{dx}{x+5}$.

ANSWER. Try and choose $u = x + 5$, then $\frac{du}{dx} = 1$, or $du = dx$, and substitute u and du back into the integral,

$$\int \frac{dx}{x+5} = \int \frac{du}{u}.$$

This is the standard integral for the logarithm, hence

$$\int \frac{dx}{x+5} = \int \frac{du}{u}\Big|_{u=x+5} = \Big[\ln|u| + C\Big]_{u=x+5} = \ln|x+5| + C.$$

□

Example 67. *Compute* $\int x^2 e^{x^3} dx$.

ANSWER. Try and choose $u = x^3$, then $du = 3x^2 dx$, and

$$\int x^2 e^{x^3} dx = \int e^{x^3}\frac{1}{3}3x^2 dx = \frac{1}{3}\int e^u du\Big|_{u=x^3} = \frac{1}{3}\Big[e^u + C_0\Big]_{u=x^3} = \frac{1}{3}e^{x^3} + C.$$

Differentiation shows that indeed $[\frac{1}{3}e^{x^3} + C]' = \frac{1}{3}e^{x^3}[x^3]' = x^2 e^{x^3}$.

□

Example 68. *Compute* $\int \frac{\cos(x)}{\cos(\sin(x))^2}dx$.

ANSWER. Try and choose $u = \sin(x)$, then $du = \cos(x)dx$, and

$$\int \frac{\cos(x)}{\cos(\sin(x))^2}dx = \int \frac{1}{\cos(u)^2}du\Big|_{u=\sin(x)},$$

the latter of which we recognize to be $\tan(u) + C$, hence,

$$\int \frac{\cos(x)}{\cos(\sin(x))^2}dx = \Big[\tan(u) + C\Big]_{u=\sin(x)} = \tan(\sin(x)) + C.$$

We check that indeed, $[\tan(\sin(x)) + C]' = \tan'(\sin(x))\sin'(x) = \frac{\cos(x)}{\cos(\sin(x))^2}$.

□

Example 69. *Compute* $\int \frac{1}{\sqrt{1-x^2}}dx$.

ANSWER. As a slightly more creative use of the substitution rule, try and choose $u = \arcsin(x)$. Then $\sin(u) = x$, $dx = \cos(u)du$, and

$$\int \frac{1}{1-x^2}dx = \int \frac{1}{\sqrt{1-\sin(u)^2}}\cos(u)du\Big|_{u=\arcsin x}$$

$$= \int \frac{1}{\sqrt{\cos(u)^2}}\cos(u)du\Big|_{u=\arcsin x} = \int du\Big|_{u=\arcsin x}$$

$$= \arcsin(x) + C.$$

To check, the derivative of $\arcsin(x)$ should be computed as was done in Example 40.

□

Exercises

Exercise 25. *Using integration by substitution, determine the following integrals*

(1) $\int 3x^2(x^3 + 6)dx$, *(do this one two times, using two different methods)*

(2) $\int 3x^2 e^{x^3+6}dx$

(3) $\displaystyle\int 2x\sqrt{1+x^2}\,dx$

(4) $\displaystyle\int \frac{(\ln x)^2}{x}\,dx$

(5) $\displaystyle\int \frac{x}{\sqrt{x^2-1}}\,dx$

(6) $\displaystyle\int \frac{3x^2+4x-4}{2x^3+4x^2-8x+1}\,dx$

(7) $\displaystyle\int x\sin(x^2)\,dx$

(8) $\displaystyle\int \tan(x)\,dx$

(9) $\displaystyle\int \sqrt{1-x^2}\,dx$

2. Integration by parts

Just as the chain rule of differentiation led to the technique of integration by substitution, the product rule of differentiation also leads to a technique for finding primitives. The product rule reads

$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x).$$

Hence, again by the definition of the indefinite integral, the primitive of $[f(x)g(x)]'$, $f(x)g(x)$, can be expressed as

$$f(x)g(x) = \int \left(f'(x)g(x) + f(x)g'(x) \right) dx.$$

By the sum rule of integration, can be written in the more useful form

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int f(x)g'(x)\,dx \qquad (\textit{partial integration rule}).$$

This equation is useful, because it often turns out that the product $f(x)g'(x)$ is considerably easier to integrate than the product $f'(x)g(x)$. Mental differentiation and integration will often expose that this may indeed be the case for a given integrand. In any case, to be able to apply partial integration (or *integration by parts*), it should be possible to find a primitive of $f'(x)$.

This rule is useful in reducing powers of terms in the integrand.

Example 70. *Compute* $\displaystyle\int xe^{3x}\,dx$.

ANSWER. Define $f'(x) = e^{3x}$ and $g(x) = x$. Then $f(x) = \int e^{3x}\,dx = \frac{1}{3}e^{3x}$, $g'(x) = 1$, and

$$\int xe^{3x}\,dx = f(x)g(x) - \int f(x)g'(x)\,dx = \frac{1}{3}xe^{3x} - \int \frac{1}{3}e^{3x}\,dx.$$

The integral in the last term is a standard integral, and hence,

$$\int xe^{3x}\,dx = \frac{1}{3}xe^{3x} - \frac{1}{3}\frac{1}{3}e^{3x} + C = \frac{1}{3}\left(x - \frac{1}{3}\right)e^{3x} + C.$$

To verify that this is a primitive of the integrand, differentiate, $[\frac{1}{3}(x - \frac{1}{3})e^{3x} + C]' = \frac{1}{3}\left([x - \frac{1}{3}]'e^{3x} + (x - \frac{1}{3})[e^{3x}]'\right) = \frac{1}{3}\left(e^{3x} + 3(x - \frac{1}{3})e^{3x}\right) = xe^{3x}$, which is indeed the integrand. $\square$

Example 71. *Compute* $\displaystyle\int x^2 e^x\,dx$

ANSWER. Define $f'(x) = e^x$, $g(x) = x^2$. Then $f(x) = e^x$, $g'(x) = 2x$, and

$$\int x^2 e^x dx = x^2 e^x - 2 \int x e^x dx.$$

The power of x in the integral in the latter term is reduced by one and is thereby simplified. We still can't solve it however, although we can reduce the power in x further, by a second round of partial integration:

$$\int x e^x dx = x e^x - \int e^x dx = x e^x - e^x = (x-1)e^x.$$

Therefore, we have

$$\int x^2 e^x dx = x^2 e^x - 2(x-1)e^x = (x^2 - 2x + 2)e^x.$$

To verify, $[(x^2 - 2x + 2)e^x]' = [(x^2 - 2x + 2)]'e^x + (x^2 - 2x + 2)[e^x]' = (2x - 2)e^x + (x^2 - 2x + 2)e^x = x^2 e^x$, which is indeed the integrand.

Example 72. *Find* $\int \ln(x)dx$.

ANSWER. Define $f'(x) = 1$, $g(x) = \ln(x)$. Then $f(x) = x$, $g'(x) = \dfrac{1}{x}$, and

$$\int \ln(x)dx = f(x)g(x) - \int f(x)g'(x)dx = x\ln(x) - \int x \cdot \frac{1}{x}dx = x\ln(x) - x + C. \qquad \square$$

Because integration is a heuristic search process, even when the rules are employed, it is always important to verify the obtained primitive through differentiation. In fact, the only way to prove that a function is a primitive through differentiation.

Exercises

Exercise 26. *Using integration by parts, find the following integrals*

(1) $\displaystyle\int x \sin(x)dx$,

(2) $\displaystyle\int x^2 \cos(x)dx$

(3) $\displaystyle\int x^2 \sqrt{1+x}dx$

(4) $\displaystyle\int \frac{xe^x}{(x+1)^2}dx$

(5) $\displaystyle\int \frac{x}{\sqrt{x+1}}dx$

Example 73. *Determine* $\int \ln\sqrt{x+1}dx$.

ANSWER. Note that $\int \ln\sqrt{x+1}dx = \frac{1}{2}\int \ln(x+1)dx$. First apply integration by substitution. Try and choose $u = x+1$, then $du = dx$, and

$$\frac{1}{2}\int \ln(x+1)dx = \frac{1}{2}\int \ln(u)du\Big|_{u=x+1}.$$

Next, apply integration by parts,

$$\int \ln(u)du = u\ln(u) - \int u \cdot \frac{1}{u}du = u\ln(u) - u + C.$$

Substituting back $u = x + 1$, we have
$$\frac{1}{2}\int \ln(x+1)dx = \frac{1}{2}(x+1)(\ln(x+1) - 1) + C.$$
To verify its correctness,
$$[(x+1)(\ln(x+1) - 1) + C]' = \frac{1}{2}(x+1)'\left(\ln(x+1) - 1\right) + \frac{1}{2}(x+1)\left(\ln(x+1) - 1\right)'$$
$$= \frac{1}{2}\left(\ln(x+1) - 1\right) + \frac{1}{2}(x+1)\frac{1}{x+1} = \ln\sqrt{x+1},$$
which is the integrand.   □

Exercise 27. *Determine the following primitives*

(1) $\displaystyle\int x(x+5)^4 dx$

(2) $\displaystyle\int x(x^2+5)^4 dx$

(3) $\displaystyle\int (3x+1)e^{x/3} dx$

(4) $\displaystyle\int x\sqrt{16 - 3x}\,dx$

(5) $\displaystyle\int \sin(x)\cos(\cos(x))dx$

3. Partial fraction expansion

We skip this technique.

4. Improper integrals

Sometimes, or rather, often in statistics, we wish to compute the area underneath a curve on an interval of infinite extend. That the area under a curve could be finite should actually be a bit surprising, but as we shall see, the fundamental theorem of calculus implies that it's true.

Example 74. *The proportion of light bulbs that still burns after an amount t time units is often accurately predicted by the area under the curve of function*
$$f(t) = \begin{cases} e^{-t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases},$$
on the interval $[0, t)$. What is the proportion of light bulbs that still work after T time units?

SOLUTION. The proportion of light bulbs that still works after t units of time is given by the area underneath the curve of $f(x)$ on the interval $[T, \infty)$—i.e., the shaded region in the figure below. That is, we have to calculate
$$\text{``}\int_T^\infty f(t)dt\text{ ''},$$
if there is such a thing. Fortunately the Fundamental Theorem of Calculus allows for a natural definition of this integral. According to the theorem,
$$\int_T^b f(t)dt = \left[\int e^{-t}dt\right]_T^b = \left[\frac{1}{-1}e^{-t}\right]_T^b = (-e^{-b}) - (-e^{-T}) = e^{-T} - e^{-b}.$$

Now, if we let b increase indefinitely, we the change in area becomes more and more negligible, and we get closer and closer to the right proportion of light bulbs. That is, if we let $b \to \infty$, then, in the limit

$$\lim_{b \to \infty} \int_T^b f(t)dt = \lim_{b \to \infty} (e^{-T} - e^{-b}) = e^{-T} - 0 = e^{-T},$$

is the proportion of light bulbs that still work after T time units.

$$f(t) = \begin{cases} e^{-t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

□
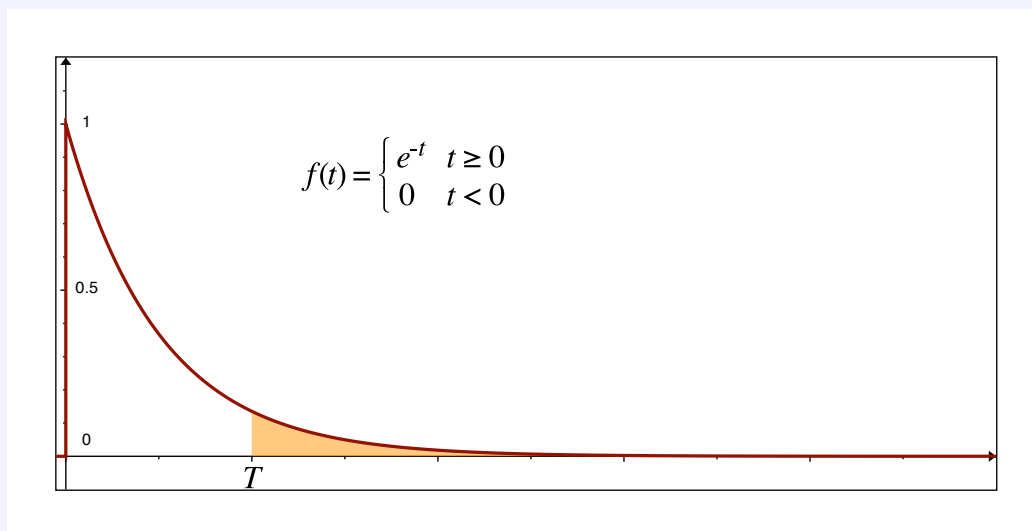
Example 74 motivates the definition of the symbols

$$\int_a^\infty f(x)dx, \qquad \int_{-\infty}^a f(x)dx, \qquad \text{and} \qquad \int_{-\infty}^\infty f(x)dx,$$

respectively as,

$$\int_a^\infty f(x)dx = \lim_{b \to \infty} \left[\int f(x)dx \right]_a^b = \lim_{b \to \infty} \int_a^b f(x)dx,$$

$$\int_{-\infty}^a f(x)dx = \lim_{a \to -\infty} \left[\int f(x)dx \right]_a^b = \lim_{a \to -\infty} \int_a^b f(x)dx,$$

$$\int_{-\infty}^\infty f(x)dx = \lim_{\substack{a \to -\infty \\ b \to \infty}} \left[\int f(x)dx \right]_a^b = \lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b f(x)dx,$$

if these limits exist. (The last of these is ambiguous, but this is not a problem if the limit is finite.) These integrals are called *improper integrals*. Note that these symbols are defined in terms of limiting values of *primitives*, and not in terms of definite integrals. In fact, it would be impossible to approximate the area underneath the curve of $f(x)$ in Example 74 by Riemann sums. The limits do not necessarily have to be $\pm\infty$; the integral

$$\lim_{a \to 0} \int_a^b \frac{1}{\sqrt{x}}dx = 2\sqrt{b} - 2\sqrt{a} = 2\sqrt{b},$$

where the limit was taken from above, is also improper, because it has a vertical asymptote at $x = 0$, as may be seen from Figure 2.
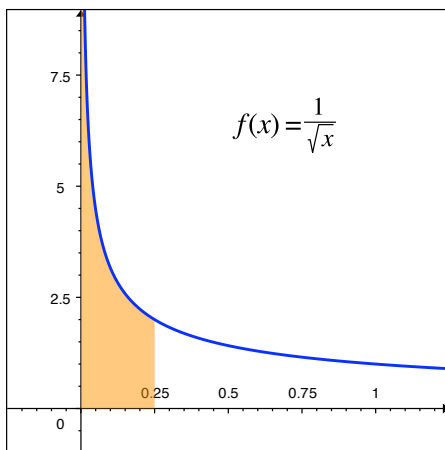
FIGURE 1. Improper integral due to vertical asymptote.

Convergence. Of course, not all functions have finite areas under the curve on an unbounded range of integration. An improper integral of the form $\int_a^\infty f(x)dx$, is said to be *convergent*, if its corresponding limit is finite,

$$\lim_{b\to\infty} \int_a^b f(x)dx = L < \infty.$$

If an improper integral is not convergent, it is called *divergent*.

Example 75. *Check whether the integral $\int_1^\infty \frac{1}{x}dx$ converges or diverges.*

ANSWER. The improper integral can be computed from the standard integral for the natural logarithm

$$\lim_{b\to\infty} \int_1^b \frac{1}{x}dx = \lim_{b\to\infty} \left[\ln|x|\right]_1^b = \lim_{b\to\infty} \left(\ln|b| - \ln|1|\right) = \lim_{b\to\infty} \ln|b| = \infty.$$

Hence, the integral diverges.   □

Example 76. *Check whether the integral $\int_1^\infty \frac{1}{x^2}dx$ converges or diverges.*

ANSWER. The improper integral can be computed from the power rule for integrals

$$\lim_{b\to\infty} \int_1^b \frac{1}{x^2}dx = \lim_{b\to\infty} \left[-\frac{1}{x}\right]_1^b = \lim_{b\to\infty} \left(\frac{-1}{b} - \frac{-1}{1}\right) = \lim_{b\to\infty} \left(1 - \frac{1}{b}\right) = 1.$$

Hence, the integral converges.   □

Care needs to be taken in evaluating improper integrals however, as the integrand may have points at which it is not defined, or where the integrand becomes unbounded. In such points the integrand may or may not converge.

Example 77. *Determine if the integral $\int_1^\infty \frac{1}{(x-3)^2} dx$ converges or diverges.*

ANSWER. If we proceded as in the previous example, the improper integral would be computed as

$$\lim_{b\to\infty} \int_1^b \frac{1}{(x-3)^2} dx = \lim_{b\to\infty} \left(\frac{-1}{b-3} - \frac{-1}{1-3}\right) = \lim_{b\to\infty} \left(-\frac{1}{b-3} - \frac{1}{2}\right) = -\frac{1}{2}. \qquad \text{wrong!}$$

This is clearly wrong because not only because the answer is negative whereas the integrand is always positive, but also because of the following: Notice the special point $x = 3$, where the integrand is not defined. If we break up the range of integration into the intervals $(1, 3), (3, 10)$, and $(10, \infty)$, then

$$\lim_{b\to\infty} \int_1^b \frac{1}{(x-3)^2} dx = \lim_{a\to 3} \int_1^a \frac{1}{(x-3)^2} dx + \lim_{b\to 3} \int_b^{10} \frac{1}{(x-3)^2} dx + \lim_{c\to\infty} \int_{10}^c \frac{1}{(x-3)^2} dx.$$

While the last improper integrals converges, the first two improper integrals both diverge—e.g., $\lim_{a\to 3} \int_1^a \frac{1}{(x-3)^2} dx = \lim_{a\to 3} \left(-\frac{1}{a-3} - \frac{1}{2}\right) = \infty$. $\qquad \square$

Improper integrals are encountered quite frequently in mathematical statistics.

Example 78. *A probability density function is a non-negative function $f(x)$ (i.e., $f(x) \geq 0$ for all $x \in \mathbb{R}$), such that*

$$\int_{-\infty}^\infty f(x)dx = 1.$$

Let $g(x) = Ce^{-3x}$ if $x \geq 0$ and $g(x) = 0$ if $x < 0$. Determine C, such that $g(x)$ is a probability density function.

ANSWER. If $g(x)$ is to be a probability density function, it must be that

$$\int_{-\infty}^\infty g(x)dx = \int_{-\infty}^0 g(x)dx + \int_0^\infty g(x)dx = \int_{-\infty}^0 0 \cdot dx + \int_0^\infty Ce^{-3x}dx = 1.$$

The last integral can be computed as

$$\lim_{b\to\infty} C \int_0^b e^{-3x}dx = \lim_{b\to\infty} C \left[\frac{1}{-3}e^{-3x}dx\right]_0^b = \lim_{b\to\infty} C \left(\frac{1}{3}e^{-3\cdot 0} - \frac{1}{3}e^{-3b}\right) = \frac{C}{3}(1-0),$$

hence

$$\int_{-\infty}^\infty g(x)dx = \frac{C}{3} = 1 \iff C = 3.$$

$\qquad \square$

Exercise 28. *Determine the ranges of $\alpha \in \mathbb{R}$ for which the improper integral*

$$\int_1^\infty \frac{1}{t^\alpha} dt$$

converges, and for which values of α it diverges.

Exercises

Exercise 29. *Determine whether and when the following improper integrals whenever are convergent, and if so, determine their value.*

(1) $\int_1^\infty \frac{1}{x^3} dx$

(2) $\displaystyle\int_0^\infty e^{-ax}dx$

(3) $\displaystyle\int_2^\infty \frac{8}{(x-5)^2}dx$

(4) $\displaystyle\int_3^\infty \frac{x^2}{\sqrt{x^3-1}}dx$

5. Numerical integration

As alluded to at the beginning of this chapter, not always does there exist a closed form expression for an indefinite integral (i.e., for the primitive of the integrand).

Example 79. *A notorious example is the Gaussian probability distribution function, which is defined as*

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du,$$

and plays a pivotal role in statistics.

Note that the mere fact that we, poor soles, cannot find a closed form expression for a given indefinite integral, doesn't mean there isn't one. Maybe we just haven't been smart enough. In such cases one might seek refuge to computer algebra systems such as Maxima, or the website Wolframalpha.com.

If none exists, thanks to the Fundamental Theorem, we still may numerically approximate its value up to a constant of integration, using Riemann sums: By the very definition of the definite integral, we subdivide the interval of integration into n segments, and evaluate $f(x)$ in points $x_1, x_2, \ldots, x_n$ inside each segment, and compute the integral as

$$\int_a^b f(x)dx \approx f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{i=1}^n f(x_i)\Delta x = M,$$

where $\Delta x = \frac{b-a}{n}$. If the x_i are taken to be exactly the point in the middle of the segment, i.e., if

$$x_i = \frac{a_i + a_{i+1}}{2},$$

where a_i is the lower limit of the segment and a_{i+1} the upper limit, then this approximation is known as the *midpoint rule* approximation.

In principle, the approximation by the midpoint rule can be made accurate to any desired degree, by taking the number of rectangles n large enough. Alternatively, instead of rectangles, one can take secant lines to approximate the curve on a segment and take area underneath the secant line to approximate the area on that segment. If a_i and a_{i+1} are the lower and upper limits of a segment, respectively, then the area under the secant line is given by

$$\text{Area}_i = f(a_i)\Delta x + \frac{1}{2}[f(a_{i+1}) - f(a_i)]\Delta x = \frac{f(a_i) + f(a_{i+1})}{2}\Delta x.$$

The area on (a, b) is then approximated by the sum of the secant areas,

$$\int_a^b f(x)dx \approx \sum_{i=1}^n \text{Area}_i = \left[\frac{1}{2}f(a_1) + f(a_2) + \cdots + f(a_n) + \frac{1}{2}f(a_{n+1})\right]\Delta x = T.$$

This method of numerical approximation is called the *Trapezoidal rule*, for obvious reasons. It can be shown that the error of approximation for the Midpoint rule is smaller than $C\frac{(b-a)^3}{24n^2}$, where

$C = \max_{x \in [a,b)} |f''(x)|$, that is

$$\left| \int_a^b f(x)dx - M \right| \le C \frac{(b-a)^2}{24n^2}.$$

For the Trapezoidal rule, this upper bound on the error of approximation is

$$\left| \int_a^b f(x)dx - T \right| \le C \frac{(b-a)^3}{12n^2},$$

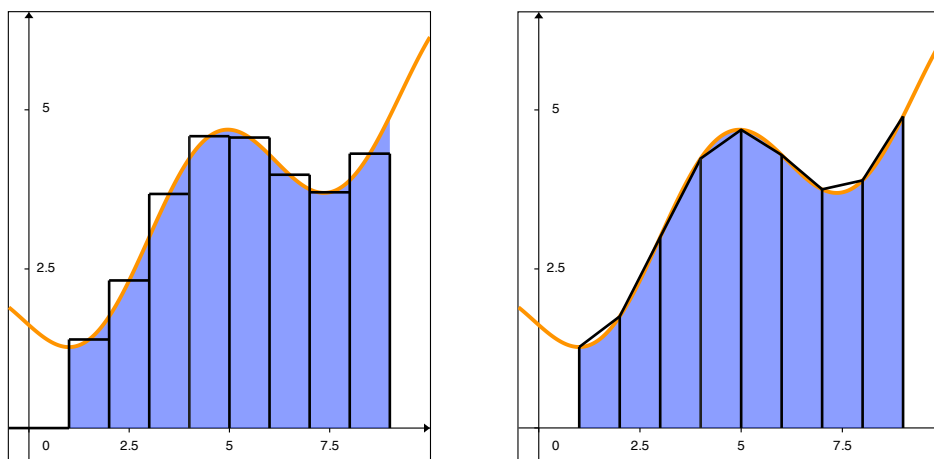which is more accurate by a factor of $1/12$ than the Midpoint rule.

FIGURE 2. Numerical approximation of integral with Midpoint rule (left) and Trapezoidal rule (left).

The essential difference between the Midpoint rule and Trapezoidal rule is that the Midpoint rule approximates the function $f(x)$ by a function that is constant on each segment, whereas the Trapezoidal rule approximates $f(x)$ by a separate linear function for each segment. An extension to the latter idea is to approximate $f(x)$ on each segment by a quadratic function of the form

$$f(x) \approx A_i x^2 + B_i x + C_i, \qquad x \in [a_i, b_i).$$

Obviously this allows for the approximation to capture some of the curvature of $f(x)$, but it does require us to determine the constants A_i, B_i, C_i on each interval $[a_i, a_{i+1})$. Fortunately there a shortcut for approximating the area on the interval (a, b) in this way, that uses the computations from the Midpoint rule and the Trapezoidal rule. If M is the approximate area obtained with the Midpoint rule, and T is the approximate area as obtained with the Trapezoidal rule, then

$$S = \frac{2}{3}M + \frac{1}{3}T$$

is the area that would be obtained with the quadratic approximation of $f(x)$ on each segment. This rule is called *Simpson's rule*, and can give a much better approximation to the area under the curve of $f(x)$ then either the Midpoint rule of the Trapezoidal rule alone. The upper bound on the error of approximation for Simpson's rule is

$$\left| \int_a^b f(x)dx - S \right| \le K \frac{(b-a)^5}{2880n^4}$$

where $K = \max_{x \in [a,b)} f''''(x)$. This is much smaller than you might have expected from considering the errors of approximation of the Midpoint and Trapezoidal rules.

Exercises

Exercise 30. *Numerically approximate the following integrals using Simpson's rule, using* $n = 4$. *Give the value up to five decimal places.*

(1) $\displaystyle\int_0^2 \sqrt{1 + 2x^3}\,dx$

(2) $\displaystyle\int_0^2 \sqrt{\sin(x)}\,dx$

(3) $\displaystyle\int_0^1 \frac{1}{2 + x^3}\,dx$

Chapter 6

Approximation and Taylor Series

1. Approximation of functions with Taylor polynomials

We motivated the definition of the derivative of functions as a limit of the secant approximation,

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a),$$

and noted that in the neighborhood of a given point a a function can be approximated by the tangent line to the graph of the function given by

$$y = p_1(x) = f(a) + f'(a)(x - a).$$

Notice, that $p_1(x)$ has the same derivative as $f(x)$ in the point a, for

$$f(x) \approx p_1'(x) = [f(a) + f'(a)(x - a)]' = f'(a)[x - a]' = f'(a),$$

and so $p_1(a) = f'(a)$.

In the neighborhood of a—that is, close to a—this approximation can be quite accurate. However, this really depends on how fast the slope $f'(a)$ of the tangent lines change as x is moved away from a. That is, the accuracy depends on $f''(a)$. Hence, one should expect a better approximation to $f(x)$ in the neighborhood of a when not only the first derivative of the approximating function is equal to the first derivative of $f(x)$ in a, but also the second order derivative is. This can be accomplished by approximating $f(x)$ with a second order polynomial

$$f(x) \approx p_2(x) = f(a) + f'(a)(x - a) + C_2(x - a)^2,$$

where we must choose C such that $p_2'(x) = f''(a)$, i.e.,

$$\begin{aligned} f''(a) = p_2''(x) &= [f(a) + f'(a)(x - a) + C(x - a)^2]'' \\ &= [f'(a)[x - a]' + C_2[(x - a)^2]']' = [f'(a) + 2 \cdot C_2(x - a)]' \\ &= 2 \cdot C_2. \end{aligned}$$

Hence, $C = \frac{f''(a)}{2}$. That is, the polynomial approximation to $f(x)$,

$$p_2(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2,$$

has both first and second order derivatives equal to those of $f(x)$ in the point a.

The approximation in a neighborhood of a by $p_2(x)$ is better than the approximation by $p_1(x)$, because it capture the amount of bending of the curve of $f(x)$. However, the accuracy clearly depends on how fast the bending changes as x moves away from a, and so, one should expect a more accurate approximation by a function of which the not only the first and second derivative, but of which also the third order derivative is equal to that of $f(x)$ in the point a. One can obtain this by adding a third order power of $x - a$ to $p_2(x)$,

$$p_3(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + C_3(x - a)^3,$$

in which C_3 is found in a similar way as C_2

Clearly, one can repeat this kind of reasoning ad nauseam, introducing higher and higher order terms in the approximation. So, in general, suppose that, as an approximation to $f(x)$, we choose an n-th order polynomial

$$p_n(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots + c_n(x - a)^n,$$

such that it has all its first n derivatives equal to those of $f(x)$ in the point a. For simplicity (but without loss of generality), assume for the moment that $a = 0$. Then

$$p_n(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n,$$

and

$$
\begin{aligned}
p_n(x) &= c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n && \Rightarrow p_n(0) = c_0 = f(0) \\
p_n'(x) &= 0 + c_1 + 2c_2 x + 3c_3 x^2 + \cdots + n c_n x^{n-1} && \Rightarrow p_n'(0) = c_1 = f'(0) \\
p_n^{(2)}(x) &= 0 + 0 + 2c_2 + 3 \cdot 2 c_3 x + \cdots + n(n-1)c_n x^{n-2} && \Rightarrow p_n^{(2)}(0) = 2! c_2 = f^{(2)}(0) \\
p_n^{(3)}(x) &= 0 + 0 + 0 + 3 \cdot 2 c_3 + \cdots + n(n-1)(n-2)c_n x^{n-2} && \Rightarrow p_n^{(3)}(0) = 3! c_3 = f^{(3)}(0) \\
&\vdots \\
p_n^{(n)}(x) &= 0 + 0 + 0 + 0 + \cdots + 0 + n! c_n x^{n-2} && \Rightarrow p_n''(0) = n! c_3 = f^{(n)}(0).
\end{aligned}
$$

Hence, $c_0 = f(0), c_1 = f'(0), c_2 = \frac{f''(0)}{2!}, c_3 = \frac{f^{(3)}(0)}{3!}, \ldots$ or in general, $c_k = \frac{f^{(k)}(0)}{k!}$, where $k! = k \cdot (k-1) \cdot (k-2) \cdots 3 \cdot 2 \cdot 1$ is the factorial function, and for general a:

$$c_k = \frac{f^{(k)}(a)}{k!}.$$

Such a polynomial approximation to the function $f(x)$ is called an n-th order Taylor polynomial *in the point a*, named after Brook Taylor who wrote about these approximations in his calculus textbook of 1717. He stated a proof[1] that given such an approximation $p_n(x)$, then for every x there is a number t_x that lies between a and x, such that

$$f(x) = p_n(x) + \frac{f^{(n+1)}(t_x)}{(n+1)!}(x - a)^{n+1}.$$

That is, the approximation can be made *exact* by evaluating $f^{(n+1)}(x)$ in a point t_x. of which we generally only know that it lies *in between* a and x. Stated differently, the *error* of the approximation of $f(x)$ by a Taylor polynomial is given by

$$\frac{f^{(n+1)}(t_x)}{(n+1)!}(x - a)^{n+1},$$

which is known as Lagrange's form of the remainder term. This can only hold, of course, if the $(n+1)$-th order derivative of $f(x)$ in a exists. But more stringently, it only holds when the $(n+1)$-th order derivative exists in the entire interval between x and a.

An important special case is when $n = 0$. Then, according to Taylor,

$$f(x) = f(a) + f'(t_x)(x - a),$$

for some $a < t_x < x$ (or $x < t_x < a$ if $x < a$), which is known as the *Mean Value Theorem*.

The fact that we don't know t_x, nor know how to find t_x in general, may make these facts seem useless. But the fact that we know that t_x must lie somewhere between x and a makes it

---

[1]It was not the first proof however; an earlier proof was given by James Gregory in 1671.

quite useful, because we can try to find the maximum of $|f^{(n+1)}(x)|$ on the interval between x and a to find an *upper bound* on the error of approximation.

Example 80. *For the function*
$$f(x) = e^x,$$
determine the n-th order Taylor polynomial in the point $x = 0$, determine an upper bound on the error for $x \in (0,1)$, and find n such that this error is smaller than $\frac{1}{10000}$.

ANSWER. The n-th order Taylor polynomial is quite simple, because $\frac{d}{d[} k]xe^x = e^x$ for all k, and $e^0 = 1$, hence,
$$p_n(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n.$$
The error is given by Lagrange's remainder formula
$$\frac{e^{tx}}{(n+1)!}x^{n+1},$$
which achieves a maximum when $x = 1$. Hence, since $1 = e^0 < e^{tx} < e^1 = e$, an upper bound for the error of approximation is given by $\frac{e}{(n+1)!}1^{n+1} = \frac{e}{(n+1)!}$.

This error is smaller than $\frac{1}{10000}$ if $\frac{e}{(n+1)!} < \frac{1}{10000}$ which is the case if $10000e < (n+1)! = n(n-1)\cdots 3 \cdot 2 \cdot 1 < n^n$ or if $\ln(10000e) < n\ln(n) < n^2 \Leftrightarrow \sqrt{\ln(10000)+1} < n \Leftrightarrow n \geq 4$. □

Note from Example 80 that as $n \to \infty$ the error $\frac{e^{tx}}{(n+1)!}x^{n+1}$ becomes smaller and smaller. In fact, for any (finite) value of x, $\frac{e^{tx}}{(n+1)!}x^{n+1} \to 0$ as $n \to \infty$.

Applications

Taylor polynomials have very many applications. For instance, they provide an easy method to find alternative expressions of polynomials.

Example 81. *Let $f(x) = 2x^3 - 9x^2 + 11x - 1$. Rewrite $f(x)$ in terms of powers of $x - 2$.*

SOLUTION. We approximate $f(x)$ with a Taylor polynomial in $x - 2$ of the same order as $f(x)$—in that case the "approximation" will be exact. That is,
$$p(x) = \frac{f(2)}{0!} + \frac{f^{(1)}(2)}{1!}(x-2) + \frac{f^{(2)}(2)}{2!}(x-2)^2 + \frac{f^{(3)}(2)}{3!}(x-2)^3,$$
where
$$f(2) = 2 \cdot 2^3 - 9 \cdot 2^2 + 11 \cdot 2 - 1 = 1$$
$$f^{(1)}(2) = 6 \cdot 2^2 - 18 \cdot 2 + 11 = -1$$
$$f^{(2)}(2) = 12 \cdot 2 - 18 = 6$$
$$f^{(3)}(2) = 12.$$
Hence, $p(x) = 1 + -1(x-2) + \frac{6}{2}(x-2)^2 + \frac{12}{6}(x-2)^3$, which certainly looks quite different from the original. □

Newton-Raphson. Many practical applications of calculus involve solving an equation of the form $g(x) = 0$, (i.e., finding x such that the equation is true). For example, in optimization problems, $g(x) = f'(x)$, where $f(x)$ is to be optimized. Sometimes such an equation can be solved by means of analytical methods, but more often, this is not possible and one has to resort

to numerical methods that systematically search for x. One such method is known as Newton-Raphson, which make use Taylor polynomial approximations of $g(x)$. It consists of the following steps: Suppose we have, after n try out values for x found a point x_n that lies "close" to the solution x, i.e., that x_n is an "approximate solution" for $g(x) = 0$.

(1) Approximate $g(x)$ in the point x_n with a first order Taylor polynomial; i.e.,

$$g(x) \approx p(x) = g(x_0) + g'(x_n)(x - x_n).$$

Then, because at the solution x, $g(x) = 0$, it should be the case that approximately,

$$0 \approx g(x_n) + g'(x_n)(x - x_n).$$

Solve this approximate equation for x, i.e.,

$$x \approx x_n - g(x_n)/g'(x_n).$$

(2) Define this approximate solution as the new, improved, approximation

$$x_{n+1} = x_n - g(x_n)/g'(x_n),$$

(3) Repeat steps 1 and 2 until $g(x_n) = 0$.

This process is illustrated in Figure **??**. Although the step 3 states that one must continue until $g(x_n) = 0$, this will in most cases imply an indefinite process of improvements of the last obtained approximate solution. In practice one therefore always have to be satisfied with an approximate solution, that is "close enough" to the true solution—i.e., that is accurate up to a certain number of decimal places. This is usually decided on the amount of change from one iteration to the next—i.e., on the difference between x_n and x_{n+1}.

Example 82. *Let $g(x) = x^2 - 2$; use the Newton-Raphson procedure to find a numerical approximation to x such that $g(x) = 0$.*

SOLUTION. Note that $x^2 - 2 = 0 \iff x = \pm\sqrt{2}$. For the Newton-Raphson procedure we need $g'(x) = 2x$. We start with $x = 1$, and apply the first iteration of the procedure

$$x_1 = x_0 - g(x_0)/g'(x_0) = x_0 - (x_0^2 - 2)/(2x_0) = \frac{x_0}{2} + \frac{1}{x_0} = \frac{1}{2} + \frac{1}{1} = \frac{3}{2}.$$

In the second iteration,

$$x_2 = x_1 - g(x_1)/g'(x_1) = \frac{x_1}{2} + \frac{1}{x_1} = \frac{3}{4} + \frac{2}{3} = \frac{17}{12}.$$

The third iteration gives

$$x_3 = \frac{x_2}{2} + \frac{1}{x_2} = \frac{17}{24} + \frac{12}{17} = \frac{577}{408}.$$

The fourth gives

$$x_4 = \frac{x_3}{2} + \frac{1}{x_3} = \frac{577}{816} + \frac{408}{577} = \frac{665857}{470832},$$

after which the decimal expansion, $\frac{665857}{470832} = 1.41421356237469$ doesn't change much any more. A calculator gives $\sqrt{2} \approx 1.414213562373$. $\square$

2. Series

We already saw that e^x is approximated by the n-th order Taylor polynomial in $x = 0$,

$$p_n(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{n!}x^n,$$

with error upper bound $\frac{e}{(n+1)!}$, on the interval $[0,1]$. This implies that

$$e = e^1 \approx p_n(1) = 2 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdot + \frac{1}{n!},$$

and in the limit that $n \to \infty$,

$$\sum_{k=0}^{n} \frac{1}{k!} \to e.$$

For ease of reference, let us define

$$S_1 = \sum_{k=0}^{1} \frac{1}{k!} = \frac{1}{0!} + \frac{1}{1!} = 1 + 1 = 2$$

$$S_2 = \sum_{k=0}^{2} \frac{1}{k!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} 2 \frac{1}{2}$$

$$S_3 = \sum_{k=0}^{3} \frac{1}{k!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} = 2 \frac{2}{3},$$

or in general

$$S_n = \sum_{k=0}^{n} \frac{1}{k!}.$$

These are called the *partial sums* of the sequence $1, 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \ldots = \{\frac{1}{k!}\}_{k=0}^{\infty}$. The limit $\lim_{n \to \infty} S_n$ is called the *sum of the series*

$$\sum_{k=0}^{\infty} \frac{1}{k!}.$$

In general, a *series* is a sequence of partial sums $\{S_k\}_{k=0}^{\infty}$ of the form

$$S_n = a_0 + a_1 + \cdots + a_n,$$

of the sequence $\{a_k\}_{k=0}^{\infty}$, and the *sum of the series* is

$$\lim_{n \to \infty} S_n = \sum_{k=0}^{\infty} a_k.$$

The symbol $\sum_{k=0}^{\infty} a_k$ is used to indicate both the sequence of partial sums $\{S_k\}$, as well as the sum of the series $\lim_{n \to \infty} S_n$. Of course this assumes that $\lim_{n \to \infty} S_n$ exists, which needs not be the case. Furthermore, the limit may not be finite, in which case the series is said to *diverge* (compare with improper integrals).

Example 83. *Let $S_n = \sum_{k=0}^{n} 1 = n + 1$. Then $\lim_{n+1} S_n = \lim_{n+1}(n+1) = \infty$.*

Example 84. *Let $S_n = \sum_{k=0}^{n}(-1)^k = 1 - 1 + 1 - 1 \cdots = 0$ if n is odd, and 1 if n is even. Then S_n keeps oscillating between 0 and 1, and hence, $\lim_{n+1} S_n$ does not exists.*

Example 85. *Let $S_n = \sum_{k=0}^{n} \frac{1}{2^n} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 2 - \frac{1}{2^n}$. Then $\lim_{n \to \infty} S_n = \lim_{n \to \infty} \left(2 - \frac{1}{2^n}\right) = 2$.*

The study of sequences and series is concerned with understanding when a series will converge, and when it will diverge. One particularly important example is the *geometric series*.

Example 86 (Geometric series). *Let $a, r \in \mathbb{R}$. Define*

$$S_n = \sum_{k=0}^{n} ar^n.$$

Then

$$S_n - rS_n = \sum_{k=0}^{n} ar^k - r \sum_{k=0}^{n} ar^k = a + ar + ar^2 + \cdots ar^n - ar - ar^2 - \cdots - ar^n - ar^{n+1}$$

$$= a - ar^{n+1},$$

from which

$$S_n - rS_n = (1-r)S_n = a(1 - r^{n+1}) \iff S_n = a\frac{1 - r^{n+1}}{1 - r},$$

provided $r \neq 1$. Hence, whether $\lim_{n\to\infty} S_n$ converges or diverges, depends on wether $\lim_{n\to\infty} r^{n+1}$ is finite or infinite. Clearly, if $|r| > 1$, then $\lim_{n\to\infty} r^n$ either diverges to ∞ if $r > 0$, or diverges because the limit does not exist if $r < 0$ (and hence, the sign of r^n oscillates). If $|r| < 1$ on the other hand, $\lim_{n\to\infty} r \to 0$, irrespective of the sign of r, and so

$$\sum_{k=0}^{\infty} ar^k = \lim_{n\to\infty} a\frac{1 - r^{n+1}}{1 - r} = \frac{a}{1 - r}.$$

If $r = 1$, then $S_n = \sum_{k=0}^{n} a \cdot 1^k = a(n+1)$, and hence $\lim S_n = \infty$, and so the series diverges to ∞.

Exercise 31. *Apply the geometric series to Example 85.*

3. Taylor expansions

Consider a special case of the geometric series of the previous section, in which $a = 1$ and $r = x$, then

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}, \qquad \text{if } |x| < 1.$$

The left hand side is a series, while the right hand side is a ordinary (rational) function of x. A series representation of a function like this is called a *power series*. Note that this representation is only valid for $x \in (-1, 1)$.

In general, an expression of the form

$$\sum_{k=0}^{\infty} a_k x^k,$$

is called a power series. A power series is a natural generalization of a Taylor polynomial, if we let the order n of the polynomial increase without bounds (i.e., if we let $n \to \infty$). If a function $f(x)$ has a power series representation

$$f(x) = \sum_{k=0}^{\infty} a_k x^k,$$

perhaps only for a restricted set of values of x, then this power series is called a *Taylor series*.[2] As was the case for the coefficients of the Taylor polynomial in the point $x = 0$, the coefficients

---

[2] More accurately, this special case of a power series expansion about the point $x = 0$ is called a *McLaurin series*. For a series development about an arbitrary point $x = x_0$, is called a Taylor series.

for the Taylor series are given by

$$a_k = \frac{f^{(k)}(0)}{k!},$$

and hence, the Taylor series of $f(x)$ is given by

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k, \qquad \text{for allowed values of } x.$$

It can be shown that the set of allowed values is always an interval of the form $(-R, R)$, for positive R. That is, the region for x where the representation is valid is symmetrical in the point 0. R can also be ∞, in which case the Taylor series converges for all $x \in \mathbb{R}$. Alternatively, the series may only converge for $x = 0$ (which is does always, obviously). The number R is called the *radius of convergence.*

One way to decide if a Taylor series exists for a given function, is to consider the error term of the Taylor polynomial: If

$$\lim_{n \to \infty} \frac{f^{(n+1)}(x)}{(n+1)!} x^{n+1} = 0,$$

then the series exists. We already saw for e^x for instance, that the error term

$$\lim_{n \to \infty} \frac{f^{(n+1)}(x)}{(n+1)!} x^{n+1} = \lim_{n \to \infty} \frac{e^{t_x}}{(n+1)!} x^{n+1} = 0,$$

and so, the Taylor series for e^x exists, and is given by the limit that $n \to \infty$ of the Taylor polynomial;

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k.$$

Similarly, one may derive that

$$\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}, \quad \text{and} \quad \cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}.$$

Exercise 32. *Find the Taylor series of the following functions*

(1) $f(x) = \ln(1 - x)$,

(2) $f(x) = \ln(1 + x)$,

(3) *Verify the Taylor series given for* $\sin(x)$ *and* $\cos(x)$.

Two main result for Taylor series are,

$$\frac{d}{dx} \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} a_k k x^{k-1} = \sum_{k=0}^{\infty} \frac{d}{dx} \left(a_k x^k \right) \quad \text{for all } x \text{ in the convergence region}$$

and

$$\int \left(\sum_{k=0}^{\infty} a_k x^k \right) dx = \sum_{k=0}^{\infty} \int (a_k x^k) dx \quad \text{for all } x \text{ in convergence region,}$$

both of which may seem straightforward from the sum and constant multiple rules of integration, however, one must keep in mind that here the sum is an "infinite sum" (rather a limit of partial sums) for which this is not necessarily true.

Example 87. *Determine the Taylor-series of* $\frac{1}{(1-x)^2}$

ANSWER. We already know from the geometric series that

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.$$

Differentiation of both sides gives

$$\frac{1}{(1-x)^2} = \frac{d}{dx} \sum_{k=0}^{\infty} x^k = \sum_{k=0}^{\infty} kx^{k-1}.$$

Exercise 33. *Determine the Taylor series for* $f(x) = \frac{1}{(1-x)^3}$.