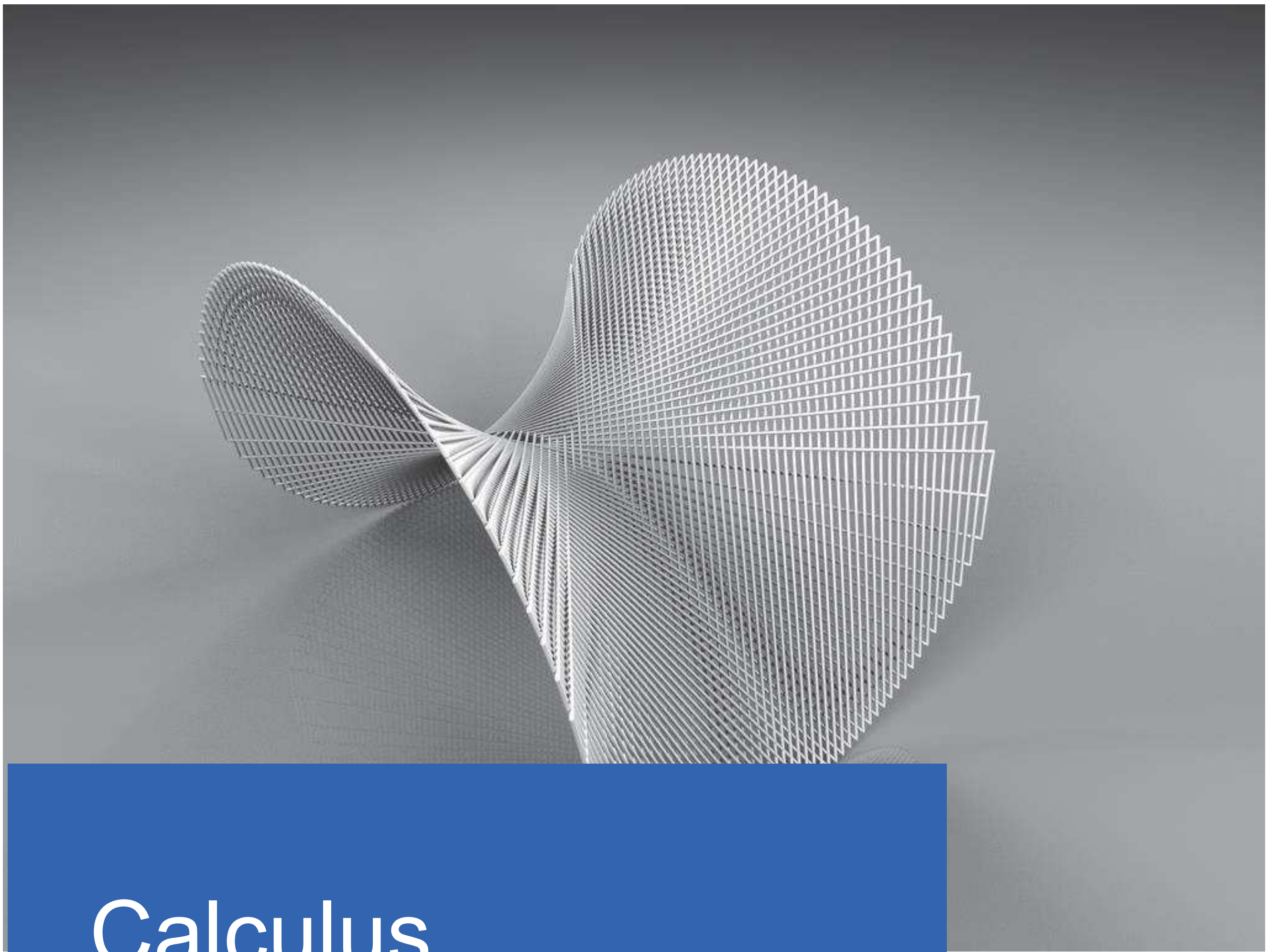
FACULTY OF
BIOSCIENCE ENGINEERING



Calculus

Prof. dr. ir. Jan Baetens
M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

Academic year 2019–2020

GHENT
UNIVERSITY

Calculus

Prof. dr. ir. Jan M. Baetens

M.Sc. Elien Van de Walle
M.Sc. An Schelfaut

The cover photo represents a hyperbolic paraboloid whose standard equation is given by

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}\,.$$

Preface

The purpose of this course is to present mathematics as the science of deductive reasoning and not as the art of manipulation. Unfortunately, many students feel mathematics is incomprehensible and is riddled with complex and abstract jargon. Our goal is to impose a lasting understanding of and appreciation for calculus on the student. Our course is intended to give the student an understanding of what calculus is truly about. It does not take more intelligence than that of a parrot to be able to go through a list of theorems and equations; but only when one understands their origins can one correctly and confidently apply them in the real world.

The over-emphasis on the calculator and foremostly the computer is definitely a point of confusion for the student. The computer is only a time-saving machine whose usefulness depends on the knowledge of the user. We do admit the computer is a remarkable machine, and we will make use of it whenever appropriate, yet it is this fascination that gives students a false sense of what they are doing. The confidence gained from all the correct answers leads to an inseparable dependence where the student is absolutely helpless without it.

Throughout the textbook we constantly refer to science and engineering. The purpose of this is to show how the scientific method applies to all disciplines and to understand that mathematics is an expression of one's observations and hypothesis. For that reason, several examples and exercises were chosen because of their relevance in reality, such that the reader can get a good feel of why and how this course is so important for future engineers. Note that because of its engineering viewpoint, we always indicate the dimensions of the used base quantities, being mass [M], time [T], temperature [Θ] and length [L]. Besides, throughout this course we include the icon in the margin of the course notes to indicate that some interactive content is available on Minerva related to the topic under discussion. At the end of every chapter one can find an extensive list of exercises linked to the topics discussed in the corresponding chapter.

Even though much time and efforts have been spent in compiling this text, it cannot be free of errors, and the authors would be grateful if these would be reported to them so that the quality of this text can be improved even further.

Finally, it goes without saying that many people have contributed to this course in addition to its authors, namely, Demir Ali Köse, Lander De Visscher, Tinne De Boeck and Ruth Van den Driessche.

Ghent, September 20, 2019

The authors

Contents

Nature laughs at the difficulties of integration.

— Pierre-Simon Laplace —

12 Integration

We have spent considerable time considering the derivatives of a function and their applications. In the following chapters, we are going to starting thinking in the other direction. That is, given a function $f(x)$, we are going to consider functions $F(x)$ such that $F'(x) = f(x)$. These functions will help us compute area, volume, mass, force, pressure, work, and much more.

12.1 Antiderivatives and (in)definite integration

12.1.1 Antiderivatives and indefinite integration

Given a function $y = f(x)$, a **differential equation** (*differentiaalvergelijking*) is one that incorporates y, x, and the derivatives of y. For instance, a simple differential equation is:

$$y' = 2x.$$

Solving a differential equation amounts to finding a function y that satisfies the given equation. Take a moment and consider that equation; can you find a function y such that $y' = 2x$?

Hopefully one was able to come up with at least one solution: $y = x^2$. Finding another may have seemed impossible until one realizes that a function like $y = x^2 + 1$ also has a derivative of $2x$. Once that discovery is made, finding yet another is not difficult; the function $y = x^2 + 123\,456\,789$ also has a derivative of $2x$. The differential equation $y' = 2x$ has many solutions. This leads us to some definitions.

Definition 12.1 (Antiderivatives and indefinite integrals)
Let a function $f(x)$ be given. An **antiderivative** (*primitieve functie*) of $f(x)$ is a function $F(x)$ such that $F'(x) = f(x)$.

The set of all antiderivatives of $f(x)$ is the **indefinite integral** (*onbepaalde integraal*) of f, denoted by

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

Note that we refer to an antiderivative of f, as opposed to the antiderivative of f, since there is always an infinite number of them. We often use upper-case letters to denote antiderivatives. Besides, knowing one antiderivative of f allows us to find infinitely more, simply by adding a constant. Not only does this give us more antiderivatives, it gives us all of them.

Theorem 12.1 (Antiderivative forms)
Let $F(x)$ and $G(x)$ be antiderivatives of $f(x)$ on an interval I. Then there exists a constant C such that, on I,

$$G(x) = F(x) + C.$$

Given a function f defined on an interval I and one of its antiderivatives F, we know all antiderivatives of f on I have the form $F(x)+C$ for some constant C. Using Definition 12.1, we can say that

$$\int f(x)\ dx = F(x) + C.$$

The integration symbol, $\int$, is in reality an elongated S, representing summing. We will later see how sums and antiderivatives are related. The function we want to find an antiderivative of is called the **integrand** (*integrandum*). It contains the differential of the variable we are integrating with respect to.

Let us now use our notice to evaluate

$$\int \sin(x)\ dx.$$

Essentially, this means that we should find all functions $F(x)$ such that $F'(x) = \sin(x)$. Of course, some thought leads us to one solution: $F(x) = -\cos(x)$, because $\frac{d}{dx}(-\cos(x)) = \sin(x)$. The indefinite integral of $\sin(x)$ is thus $-\cos(x)$, plus a constant of integration C. So:

$$\int \sin(x)\ dx = -\cos(x) + C.$$

To fully understand what is happening, it is important to realise that the process of antidifferentiation is really solving a differential question. The integral

$$\int \sin(x)\ dx$$

presents us with a differential, $dy = \sin(x)\ dx$. It is asking: What is y? We found lots of solutions, all of the form $y = -\cos(x) + C$.

Letting $dy = \sin(x)\ dx$, rewrite

$$\int \sin(x)\ dx \quad \text{as} \quad \int dy.$$

This is asking: "What functions have a differential of the form dy?" The answer is "Functions of the form $y+C$, where C is a constant." What is y? We have lots of choices, all differing by a constant; the simplest choice is $y = -\cos(x)$.

In Mathematica, we can use the command **Integrate** to evaluate an indefinite integral. For instance,

$$\int (3x^2 + 4x + 5)\ dx.$$

can be evaluated as follows.

```
In[19]:= Integrate[3*x^2+4*x+5, x]

Out[19]=  5x +2x^2 +x^3
```

We can also see that taking the derivative of our answer returns the function in the integrand. Thus we can say that:

$$\frac{d}{dx}\left(\int f(x)\ dx\right) = f(x).$$

Differentiation undoes the work done by antidifferentiation.

Taking into account the lists of derivatives of algebraic and transcendental functions presented in Chapter 9, we may now state some important antiderivatives. We easily see that

$$\int 0\ dx = C\,,$$

and

$$\int 1\ dx = \int dx = x + C\,,$$

from which we can infer the following more general integral rule:

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

for $n \neq -1$.

For what concerns the exponential and logarithmic functions, we get the following derivative functions:

- $\int e^x\ dx = e^x + C,$
- $\int a^x\ dx = \frac{1}{\ln(a)}a^x + C,$
- $\int \frac{1}{x}\ dx = \ln|x| + C,$

while for the trigonometric and hyperbolic functions we get:

- $\int \sin(x)\ dx = -\cos(x) + C$
- $\int \cos(x)\ dx = \sin(x) + C$
- $\int \sec^2(x)\ dx = \tan(x) + C$
- $\int \csc^2(x)\ dx = -\cot(x) + C$
- $\int \sinh(x)\ dx = \cosh(x) + C$
- $\int \cosh(x)\ dx = \sinh(x) + C$
- $\int \frac{1}{\cosh^2(x)}\ dx = \tanh(x) + C$
- $\int \frac{1}{\sinh^2(x)}\ dx = -\coth(x) + C$

Besides, we have the following properties, which are completely in line with those for derivatives (Theorem 9.3)

Theorem 12.2 (Properties of the antiderivative)
Let f and g be differentiable on an open interval I and let k be a real number. Then:

1. *Sum/Difference rule:*

$$\int \big(f(x) \pm g(x)\big)\, dx = \int f(x)\, dx \pm \int g(x)\, dx\,. \tag{12.1}$$

2. *Constant multiple rule:*

$$\int k f(x)\, dx = k \int f(x)\, dx\,. \tag{12.2}$$

For the sake of illustration, We will prove the sum rule. The proofs of the other properties proceed in a similar way.

Suppose that $F(x)$ is an anti-derivative of $f(x)$ and that $G(x)$ is an anti-derivative of $g(x)$. So we have that $F'(x) = f(x)$ and $G'(x) = g(x)$. Basic properties of derivatives also tell us that

$$\big(F(x) + G(x)\big)' = F'(x) + G'(x) = f(x) + g(x)\,,$$

and so $F(x) + G(x)$ is an anti-derivative of $f(x) + g(x)$. In other words,

$$\int f(x) + g(x)\, dx = F(x) + G(x) + C = \int f(x)\, dx + \int g(x)\, dx\,.$$

In Section 9.1.4 we saw that the derivative of a position function gave a velocity function, and the derivative of a velocity function describes acceleration. We can now go the other way: the antiderivative of an acceleration function gives a velocity function, etc. While there is just one derivative of a given function, there are infinitely many antiderivatives. Therefore we cannot ask "What is the velocity of an object whose acceleration is -32m/s^2?", since there is more than one answer.

We can find the answer if we provide more information with the question, as done in the following example. Often the additional information comes in the form of an initial value, a value of the function that one knows beforehand.

Example 12.1

The acceleration due to gravity of a falling object is $-9\ \text{m/s}^2$. At time $t = 3$, a falling object had a velocity of -10 m/s. Find the equation of the object's velocity.

Solution

We want to know a velocity function, $v(t)$. We know two things:

- The acceleration, i.e., $v'(t) = -9$, and
- the velocity at a specific time, i.e., $v(3) = -10$.

Using the first piece of information, we know that $v(t)$ is an antiderivative of $v'(t) = -9$. So we begin by finding the indefinite integral of -9:

$$\int v'(t)\ dt = \int (-9)\ dt = -9t + C = v(t).$$

Now we use the fact that $v(3) = -10$ by plugging in this point in the equation we just got for $v(t)$:

$$-9 \cdot (3) + C = -10\,,$$

for which it directly follows that $C = 37$.

Thus $v(t) = -9t + 37$. We can use this equation to understand the motion of the object: when $t = 0$, the object had a velocity of $v(0) = 37$ m/s. Since the velocity is positive, the object was moving upward.

In the remainder of this section, we will see how position and velocity are unexpectedly related by the areas of certain regions on a graph of the velocity function.

12.1.2 The definite integral

We start with an easy problem. An object travels in a straight line at a constant velocity of 5m/s for 10 seconds. How far away from its starting point is the object?

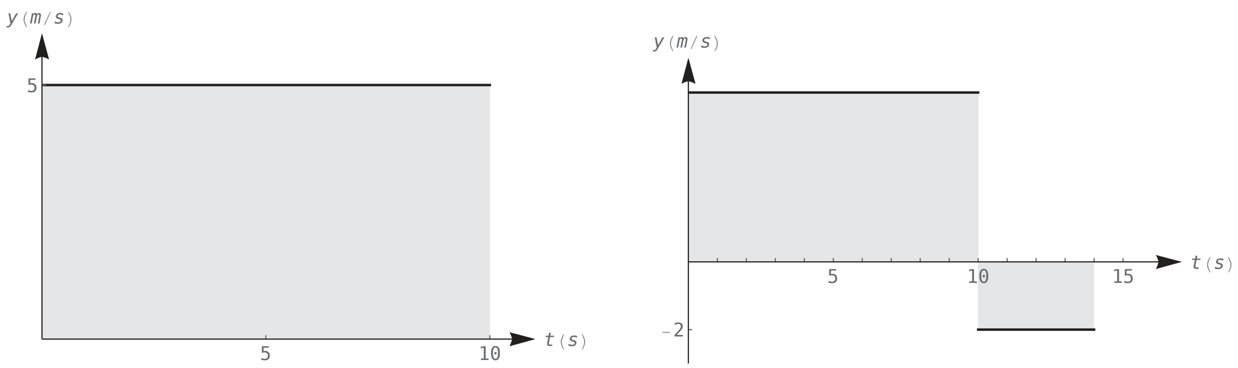
Since, we have that Distance = Rate × Time, it follows that this distance is 50 metres. This solution can be represented graphically. Consider Figure 12.1(a), where the constant velocity of 5m/s is graphed on the axes. Shading the area under the line from $t = 0$ to $t = 10$ gives a rectangle with an area of 50 square units; when one considers the units of the axes, we can say this area represents 50 m.

Now consider a slightly harder situation (and not particularly realistic): an object travels in a straight line with a constant velocity of 5m/s for 10 seconds, then instantly reverses course at a rate of 2m/s for 4 seconds. How far away from the starting point is the object – what is its displacement?

Here, we get:

$$\text{Distance} \ = 5 \cdot 10 + (-2) \cdot 4 = 42 \text{ m}.$$

Hence the object is 42 metres from its starting location.



(a) (b)

Figure 12.1: The total displacement of an object travelling in a straight line at a constant velocity of 5m/s for 10 seconds (a) and an object travelling a straight line with a constant velocity of 5m/s for 10 seconds, and then instantly reversing course at a rate of 2m/s for 4 seconds (b).

We can again depict this situation graphically. In Figure 12.1(b) we have the velocities graphed as straight lines on $[0,10]$ and $[10,14]$, respectively. The displacement of the object is given by

$$\text{Area above the } t\text{–axis} \quad - \quad \text{Area below the } t\text{–axis},$$

which is easy to calculate as $50-8=42$ metres.

These examples do not prove a relationship between area under a velocity function and displacement, but it does imply a relationship exists. Section 12.3 will fully establish fact that the area under a velocity function is displacement.

Anyhow, given a graph of a function $y=f(x)$, we will find that there is great use in computing the area between the curve $y=f(x)$ and the x-axis. Because of this, we need to define some terms.

Definition 12.2 (The definite integral, total signed area)
Let $y=f(x)$ be defined on a closed interval $[a,b]$. The total signed area from $x=a$ to $x=b$ between f and the x-axis is:

$$(\text{area under } f \text{ and above the } x\text{–axis on } [a,b]) - (\text{area above } f \text{ and under the } x\text{–axis on } [a,b]).$$

The **definite integral** (*bepaalde integraal*) of f on $[a,b]$ is the total signed area of f on $[a,b]$, denoted

$$\int_a^b f(x)\,dx,$$

where a and b are the bounds of integration.

By our definition, the definite integral gives the signed area under f. We usually drop the word signed when talking about the definite integral, and simply say the definite integral gives the area under f or, more commonly, the area under the curve. The indefinite integral and definite integral are very much related, as we will see in Section 12.3.

Let us now practice this definition.

Example 12.2

Consider the function f given in Figure 12.2.

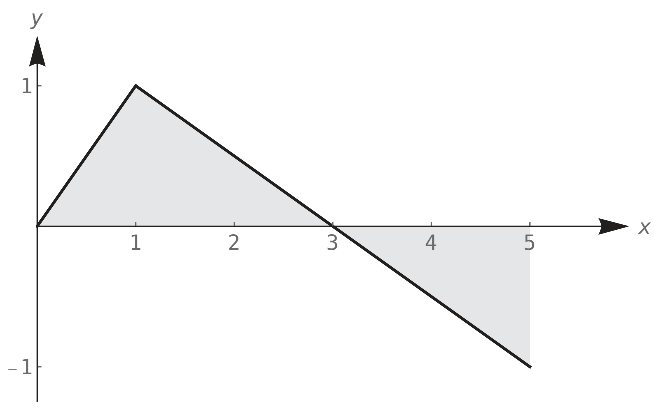


Figure 12.2: A graph of $f(x)$ in Example 12.2.

Find:

1. $\displaystyle\int_0^3 f(x)\,dx$
2. $\displaystyle\int_3^5 f(x)\,dx$
3. $\displaystyle\int_0^5 f(x)\,dx$
4. $\displaystyle\int_0^3 5f(x)\,dx$
5. $\displaystyle\int_1^1 f(x)\,dx$

Solution

1. This definite integral is the area under f on the interval $[0,3]$. This region is a triangle, so the area is
$$\int_0^3 f(x)\,dx = \frac{1}{2}(3)(1) = 1.5.$$
2. This definite integral represents the area of the triangle found under the x–axis on $[3,5]$. The area is $1/2(2)(1) = 1$; since it is found under the x–axis, this is negative area. So,
$$\int_3^5 f(x)\,dx = -1.$$
3. This definite integral is the total signed area under f on $[0,5]$. This is $1.5 + (-1) = 0.5$.
4. This definite integral is the area under $5f$ on $[0,3]$. Again, the region is a triangle, with height 5 times that of the height of the original triangle. Thus the area is
$$\int_0^3 5f(x)\,dx = \frac{1}{2}(15)(1) = 7.5.$$
5. This definite integral is the area under f on the interval $[1,1]$. This describes a line segment, not a region; it has no width. Therefore the area is 0.

This example illustrates some of the properties of the definite integral, listed in the following theorem.

Theorem 12.3 (Properties of the definite integral)
Let f and g be defined on a closed interval I that contains the values a, b and c, and let k be a constant. The following hold:

1. $\displaystyle\int_a^a f(x)\,dx = 0,$

2. $\displaystyle\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx,$

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

4. $\displaystyle\int_a^b \big(f(x) \pm g(x)\big)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx,$

5. $\displaystyle\int_a^b k f(x)\,dx = k \cdot \int_a^b f(x)\,dx.$

The proofs of these properties will be provided at the end of the next section once we have a better understanding of definite integrals through the conceptualisation of Riemann sums.

The area definition of the definite integral allows us to use geometry to compute the definite integral of some simple functions.

Example 12.3

Evaluate the following definite integrals:

1. $\displaystyle\int_{-2}^{5} (2x-4)\,dx$
2. $\displaystyle\int_{-3}^{3} \sqrt{9-x^2}\,dx.$

Solution

1. It is useful to sketch the function in the integrand, as shown in Figure 12.3(a). We see we need to compute the areas of two regions, which we have labelled R_1 and R_2. Both are triangles, so the area computation is straightforward:

$$R_1 : \frac{1}{2}(4)(8) = 16 \qquad R_2 : \frac{1}{2}(3)6 = 9.$$

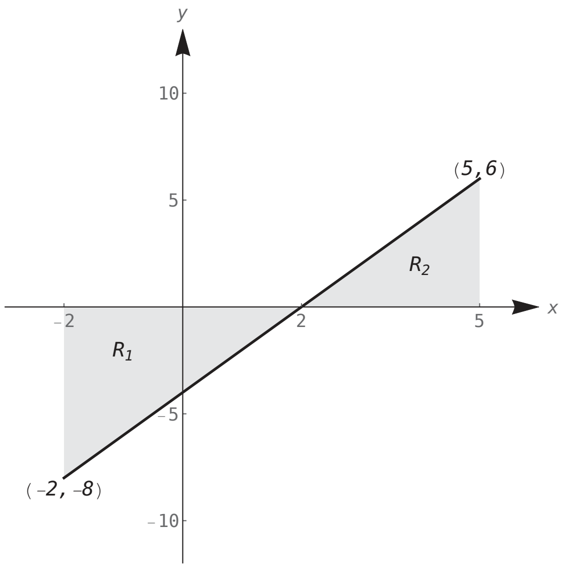
Region R_1 lies under the x–axis, hence it is counted as negative area, so

$$\int_{-2}^{5} (2x-4)\,dx = -16 + 9 = -7.$$

We may check this answer in Mathematica as follows

```
In[20]:= Integrate[2*x-4, x,-2,5]

Out[20]= -7
```



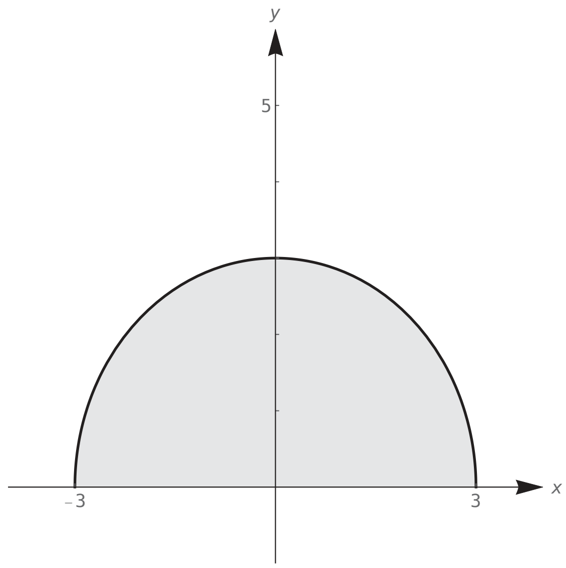


Figure 12.3: A graph of $f(x) = 2x - 4$ in (a) and $f(x) = \sqrt{9-x^2}$ in (b), from Example 12.3.

2. Recognize that the integrand of this definite integral describes a half circle, as sketched in Figure 12.3(b), with radius 3. Thus the area is:

$$\int_{-3}^{3} \sqrt{9-x^2}\, dx = \frac{1}{2}\pi r^2 = \frac{9}{2}\pi.$$

12.2 Riemann sums

In our previous examples, we have either found the areas of regions that have nice geometric shapes or the areas were given to us. But what is, for instance, the area of a region below $y = x^2$? The function $y = x^2$ is relatively simple, yet the shape it defines has an area that is not simple to find geometrically. In this section we will explore how to find the areas of such regions.

12.2.1 Approximating areas

Consider the region given in Figure 12.4, which is the area under $y = 4x - x^2$ on $[0,4]$. What is the signed area of this region – i.e., what is $\int_0^4 (4x - x^2)\, dx$? We start by approximating. We can surround the region with a rectangle with height and width of 4 and find the area is approximately 16 square units. This is obviously an over–approximation; we are including area in the rectangle that is not under the parabola.

We have an approximation of the area, using one rectangle. How can we refine our approximation to make it better? The key to this section is this answer: use more rectangles. Let us use 4 rectangles with an equal width of 1. This partitions the interval $[0,4]$ into 4 subintervals, $[0,1]$, $[1,2]$, $[2,3]$ and $[3,4]$. On each subinterval we will draw a rectangle.

There are three common ways to determine the height of these rectangles: the **left hand rule** (*linkerhand regel*), the **right hand rule** (*rechterhand regel*), and the **midpoint rule** (*midpoint regel*). The left hand rule says to evaluate the function at the left–hand endpoint of the subinterval and make the

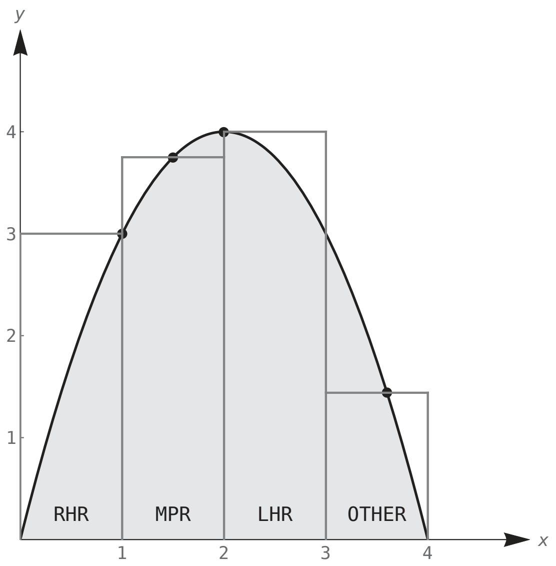


Figure 12.4: A graph of $f(x) = 4x - x^2$ and approximating $\int_0^4 (4x - x^2)\ dx$ using rectangles.

rectangle that height. In Figure 12.4, the rectangle drawn on the interval $[2, 3]$ has height determined by the left hand rule (LHR); it has a height of $f(2)$.

The right hand rule (RHR) says the opposite: on each subinterval, evaluate the function at the right endpoint and make the rectangle that height. In Figure 12.4, the rectangle drawn on $[0, 1]$ is drawn using $f(1)$ as its height. The midpoint rule (MPR) says to evaluate the function at the midpoint of each subinterval, and to make the rectangle that height. The rectangle drawn on $[1, 2]$ was made using the midpoint rule, with a height of $f(1.5)$.

These are the three most common rules for determining the heights of approximating rectangles, but one is not forced to use one of these three methods. The rectangle on $[3, 4]$ has a height of approximately $f(3.53)$, very close to the midpoint rule. It was chosen so that the area of the rectangle is exactly the area of the region under f on $[3, 4]$.

The following example will put these rules into practice.

Example 12.4

Approximate the value of

$$\int_0^4 (4x - x^2)\ dx$$

using the left hand rule, the right hand rule, and the midpoint rule, using 4 equally spaced subintervals.

Solution

We break the interval $[0, 4]$ into four subintervals as before. In Figure 12.5(a) we see 4 rectangles drawn on $f(x) = 4x - x^2$ using the left hand rule. The areas of the rectangles are given in each figure.

Note how in the first subinterval, $[0, 1]$, the rectangle has height $f(0) = 0$. We add up the areas of each rectangle (height×width) for our left hand rule approximation:

$$f(0)\cdot 1 + f(1)\cdot 1 + f(2)\cdot 1 + f(3)\cdot 1 = 0 + 3 + 4 + 3 = 10\,.$$

Figure 12.5(b) shows 4 rectangles drawn under f using the right hand rule; note how the $[3,4]$ subinterval has a rectangle of height 0.

In this example, these rectangles seem to be the mirror image of those found in Figure 12.5(a). This is because of the symmetry of our shaded region. Our approximation gives the same answer as before, though calculated a different way:

$$f(1)\cdot 1+f(2)\cdot 1+f(3)\cdot 1+f(4)\cdot 1=3+4+3+0=10\,.$$

Figure 12.5(c) shows 4 rectangles drawn under f using the midpoint rule. This gives :

$$f(0.5)\cdot 1+f(1.5)\cdot 1+f(2.5)\cdot 1+f(3.5)\cdot 1=1.75+3.75+3.75+1.75=11\,.$$

Our three methods provide two approximations, namely 10 and 11.

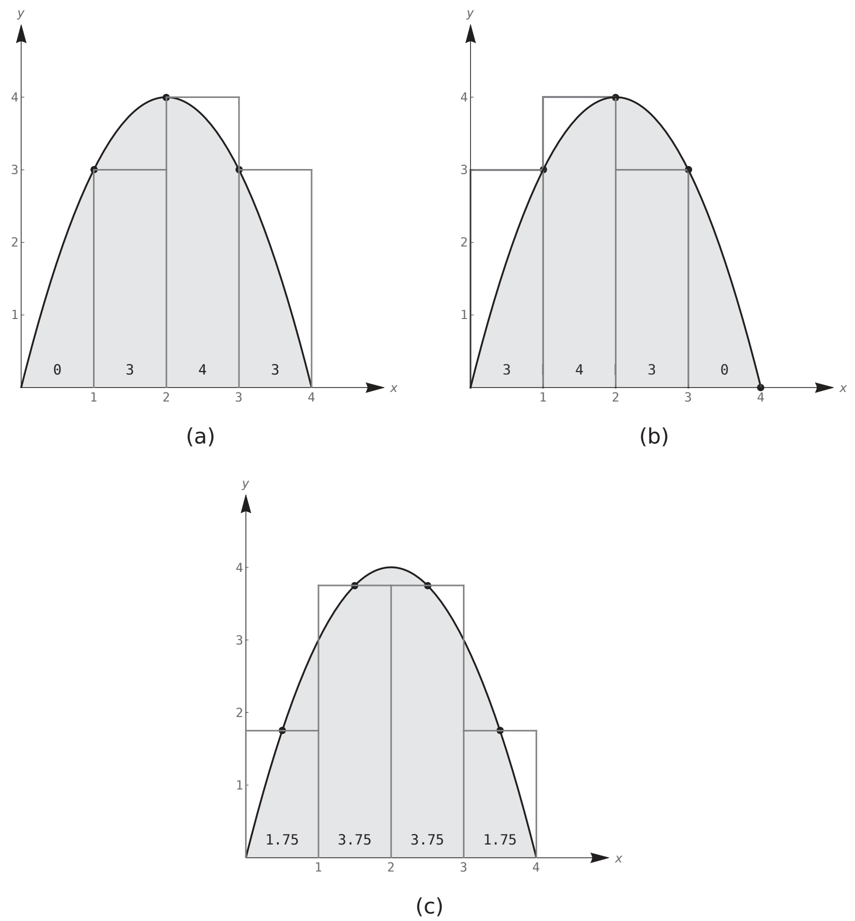


Figure 12.5: Approximating $\int_0^4 (4x-x^2)\ dx$ in Example 12.4 using the left hand rule (a), the right hand rule (b) and the midpoint rule (c).

It is hard to tell at this moment which is a better approximation. We can continue to refine our approximation by using more rectangles.

12.2.2 Riemann sums

Consider again $\int_0^4 (4x - x^2)\ dx$. We divide or partition the number line of $[0,4]$ into 16 equally spaced subintervals. We denote 0 as x_1 , so in general, we have

$$x_i = x_1 + (i-1)\Delta x\,,$$

where $i = 1, 2, \ldots, 16$ For the sake of simplicity, we will often write $\Delta x = \Delta x_i$, where Δx_i is the width of the i^{th} subinterval, whenever the width of the subintervals is the same.

Given any subdivision of $[0,4]$, the first subinterval is $[x_1, x_2]$; the second is $[x_2, x_3]$; the i^{th} subinterval is $[x_i, x_{i+1}]$. Hence, when using the left hand rule, the height of the i^{th} rectangle will be $f(x_i)$. When using the right hand rule, the height of the i^{th} rectangle will be $f(x_{i+1})$, and finally, when using the midpoint rule, the height of the i^{th} rectangle will be

$$f\left(\frac{x_i + x_{i+1}}{2}\right).$$

We illustrate this in the next example.

Example 12.5

Approximate

$$\int_0^4 (4x - x^2)\ dx$$

using the right hand rule with 16 and 1000 equally spaced intervals.

Solution

Using 16 equally spaced intervals and the right hand rule, we can approximate the definite integral as

$$\sum_{i=1}^{16} f(x_{i+1})\Delta x,$$

where we have $\Delta x = 4/16 = 0.25$. Moreover, since $x_1 = 0$, we have

$$\begin{aligned} x_{i+1} &= 0 + \big((i+1)-1\big)\Delta x \\ &= i\Delta x\,. \end{aligned}$$

Using summation formulas, we may now consider:

$$\begin{aligned} \int_0^4 (4x - x^2)\ dx &\approx \sum_{i=1}^{16} f(x_{i+1})\Delta x = \sum_{i=1}^{16} f(i\Delta x)\Delta x \\ &= \sum_{i=1}^{16} \big(4i\Delta x - (i\Delta x)^2\big)\Delta x = \sum_{i=1}^{16} (4i\Delta x^2 - i^2\Delta x^3) \\ &= (4\Delta x^2)\sum_{i=1}^{16} i - \Delta x^3 \sum_{i=1}^{16} i^2 \qquad (12.3) \\ &= (4\Delta x^2)\frac{16\cdot 17}{2} - \Delta x^3 \frac{16(17)(33)}{6} = 10.625 \qquad (\Delta x = 0.25) \end{aligned}$$

(12.4)

We were able to sum up the areas of 16 rectangles with very little computation. In Figure 12.6 the function and the 16 rectangles are graphed. While some rectangles over–approximate the area, other under–approximate the area by about the same amount. Thus our approximate area of 10.625 is likely a fairly good approximation.

For what concerns the approximation based on 1000 equally spaced, we can just use Equation (12.3); after replacing the 16's to 1000's and appropriately changing the value of Δx.

We do so here, skipping from the original summand to the equivalent of Equation (12.3) to save space. Note that $\Delta x = 4/1000 = 0.004$.

$$\begin{aligned}
\int_0^4 (4x - x^2)\, dx &\approx \sum_{i=1}^{1000} f(x_{i+1})\Delta x \\
&= (4\Delta x^2)\sum_{i=1}^{1000} i - \Delta x^3 \sum_{i=1}^{1000} i^2 \\
&= (4\Delta x^2)\frac{1000\cdot 1001}{2} - \Delta x^3 \frac{1000(1001)(2001)}{6} \\
&= 10.666656
\end{aligned}$$

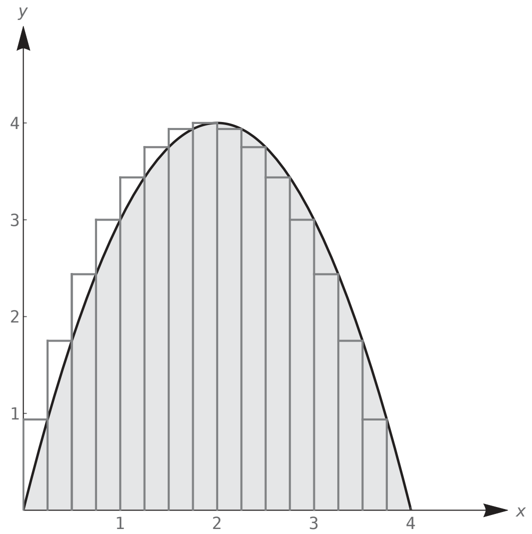


Figure 12.6: Approximating $\int_0^4 (4x - x^2)\, dx$ with the right hand rule and 16 evenly spaced subintervals.

Using many, many rectangles, we have a likely good approximation of $\int_0^4 (4x - x^2)\Delta x$. That is,

$$\int_0^4 (4x - x^2)\, dx \approx 10.666656.$$

Instead of approximating a definite integral using rectangles of the same width and height determined by evaluating f at a particular point in each consecutive subinterval, we could partition an interval $[a, b]$ with subintervals that do not have the same size. We refer to the length of the i^{th} subinterval as Δx_i. Also, one could determine each rectangle's height by evaluating f at any point c_i in the i^{th}

subinterval. Thus the height of the i^{th} subinterval would be $f(c_i)$, and the area of the i^{th} rectangle would then be $f(c_i)\Delta x_i$.

These ideas are formally defined below.

Definition 12.3 (Partition)
A **partition** (*partitie*) of a closed interval $[a,b]$ is a set of numbers $x_1, x_2, \ldots x_{n+1}$ where

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b.$$

The length of the i^{th} subinterval, $[x_i, x_{i+1}]$, is $\Delta x_i = x_{i+1} - x_i$. If $[a,b]$ is partitioned into subintervals of equal length, we let Δx_i represent the length of each subinterval.

The size of the partition, denoted $\mathcal{L}$, is the length of the largest subinterval of the partition, i.e. $\mathcal{L} = \max\limits_i (\Delta x_i)$.

Summations of rectangles with area $f(c_i)\Delta x_i$ are named after mathematician Georg Friedrich Bernhard Riemann, as given in the following definition.

Definition 12.4 (Riemann sum)
Let f be defined on a closed interval $[a,b]$, let $\{x_1, x_2, \ldots, x_{n+1}\}$ be a partition of $[a,b]$ and let c_i denote any value in the i^{th} subinterval.

The sum

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

is a **Riemann sum** (*Riemann som*) of f on $[a,b]$.

Usually Riemann sums are calculated using one of the three methods we have introduced. The uniformity of construction makes computations easier. So

$$\int_a^b f(x)\, dx$$

is typically approximated by means of the following Riemann sum

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

for which we take the following steps.

1. Divide the interval $[a,b]$ in n subintervals have equal length, such that

   $$\Delta x_i = \Delta x = \frac{b-a}{n}$$

   and the i^{th} term of the equally spaced partition is

   $$x_i = a + (i-1)\Delta x\,.$$

   Thus $x_1 = a$ and $x_{n+1} = b$.

2. Evaluate one of the following summations:

   (a) using the left hand rule we get the so-called **left Riemann sum** (*linker Riemann som*):

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

   (b) using the right hand rule we get the so-called **right Riemann sum** (*rechter Riemann som*):

$$\sum_{i=1}^{n} f(x_{i+1})\Delta x,$$

   (c) and using the midpoint rule we get the **middle Riemann sum** (*midden Riemann som*):

$$\sum_{i=1}^{n} f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x\,.$$

Example 12.6

Approximate

$$\int_{-2}^{3} (5x+2)\,dx$$

using the midpoint rule and 10 equally spaced intervals.

Solution

We have $\Delta x = 1/2$ and $x_i = (-2)+(1/2)(i-1) = i/2-5/2$ for $i = 1,2,\ldots,10$. As we are using the midpoint rule, we will also need x_{i+1} and $\frac{x_i+x_{i+1}}{2}$:

$$\frac{x_i+x_{i+1}}{2} = \frac{(i/2-5/2)+\big((i+1)/2-5/2\big)}{2} = \frac{i-9/2}{2} = \frac{i}{2}-\frac{9}{4}.$$

We now construct the Riemann sum and compute its value.

$$\begin{aligned}
\int_{-2}^{3} (5x+2)\,dx &\approx \sum_{i=1}^{10} f\left(\frac{x_i+x_{i+1}}{2}\right)\Delta x \\
&= \sum_{i=1}^{10} f\left(\frac{i}{2}-\frac{9}{4}\right)\Delta x \\
&= \sum_{i=1}^{10}\left(5\left(\frac{i}{2}-\frac{9}{4}\right)+2\right)\Delta x \\
&= \Delta x \sum_{i=1}^{10}\left[\left(\frac{5}{2}\right)i-\frac{37}{4}\right] \\
&= \Delta x\left(\frac{5}{2}\sum_{i=1}^{10}(i)-\sum_{i=1}^{10}\left(\frac{37}{4}\right)\right) \\
&= \frac{1}{2}\left(\frac{5}{2}\cdot\frac{10(11)}{2}-10\cdot\frac{37}{4}\right)
\end{aligned}$$

$$= \frac{45}{2} = 22.5$$

Note the graph of $f(x) = 5x+2$ in Figure 12.7. The regions whose area is computed by the definite integral are triangles, meaning we can find the exact answer without summation techniques. We find that the exact answer is indeed 22.5. One of the strengths of the midpoint rule is that often each rectangle includes area that should not be counted, but misses other area that should. When the partition size is small, these two amounts are about equal and these errors almost cancel each other out. In this example, since our function is a line, these errors are exactly equal and they do cancel each other out, giving us the exact answer.

Note too that when the function is negative, the rectangles have a negative height. When we compute the area of the rectangle, we use $f(c_i)\Delta x$; when f is negative, the area is counted as negative.

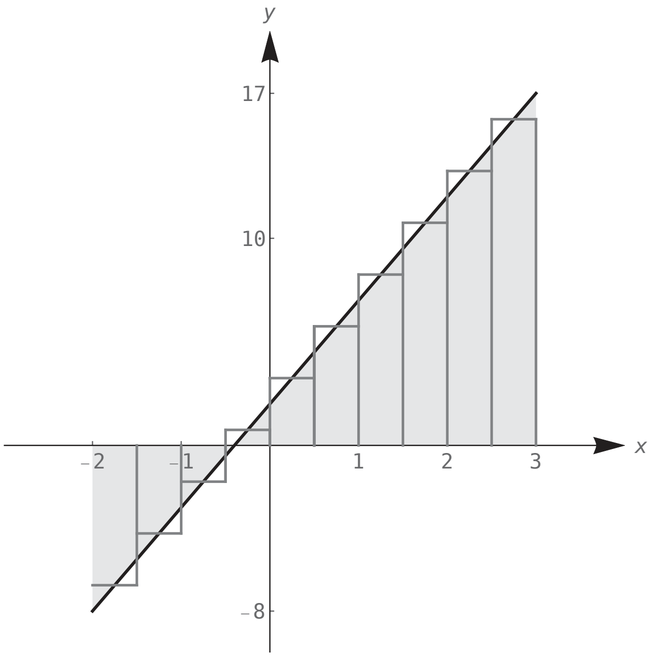


Figure 12.7: Approximating $\int_{-2}^{3}(5x+2)\ dx$ using the midpoint rule and 10 evenly spaced subintervals in Example 12.6.

Notice in the previous example that while we used 10 equally spaced intervals, this number did not play a big role in the calculations until the very end. Mathematicians love to abstract ideas; let us approximate the area of another region using n subintervals, where we do not specify a value of n until the very end.

Example 12.7

Revisit

$$\int_0^4 (4x - x^2)\ dx$$

yet again. Approximate this definite integral using the right hand rule with n equally spaced subintervals.

Solution

We know $\Delta x = (4-0)/n = 4/n$. We also find $x_i = 0+\Delta x(i-1) = 4(i-1)/n$. The right hand rule uses x_{i+1}, which is $x_{i+1} = 4i/n$.

We construct the right Riemann sum as follows.

$$\begin{aligned}
\int_0^4 (4x-x^2)\,dx &\approx \sum_{i=1}^{n} f(x_{i+1})\Delta x \\
&= \sum_{i=1}^{n} f\left(\frac{4i}{n}\right)\Delta x \\
&= \sum_{i=1}^{n}\left[4\frac{4i}{n}-\left(\frac{4i}{n}\right)^2\right]\Delta x \\
&= \sum_{i=1}^{n}\left(\frac{16\Delta x}{n}\right)i-\sum_{i=1}^{n}\left(\frac{16\Delta x}{n^2}\right)i^2 \\
&= \left(\frac{16\Delta x}{n}\right)\sum_{i=1}^{n} i-\left(\frac{16\Delta x}{n^2}\right)\sum_{i=1}^{n} i^2 \\
&= \left(\frac{16\Delta x}{n}\right)\cdot\frac{n(n+1)}{2}-\left(\frac{16\Delta x}{n^2}\right)\frac{n(n+1)(2n+1)}{6} \\
&= \frac{32(n+1)}{n}-\frac{32(n+1)(2n+1)}{3n^2} \\
&= \frac{32}{3}\left(1-\frac{1}{n^2}\right)
\end{aligned}$$

The result is an amazing, easy to use formula. To approximate the definite integral with 10 equally spaced subintervals and the right hand rule, set $n = 10$ and compute

$$\int_0^4 (4x-x^2)\,dx \approx \frac{32}{3}\left(1-\frac{1}{10^2}\right) = 10.56.$$

Recall how earlier we approximated the definite integral with 4 subintervals; with $n = 4$, the formula gives 10, our answer as before.

We now take an important leap. More precisely, for any finite n, we know that

$$\int_0^4 (4x-x^2)\,dx \approx \frac{32}{3}\left(1-\frac{1}{n^2}\right).$$

Both common sense and high–level mathematics tell us that as n gets large, the approximation gets better. In fact, if we take the limit as $n \to +\infty$, we get the exact are we are looking for, that is:

$$\int_0^4 (4x-x^2)\,dx = \lim_{n\to+\infty}\frac{32}{3}\left(1-\frac{1}{n^2}\right)$$

$$= \frac{32}{3}(1-0)$$
$$= \frac{32}{3}.$$

This is a fantastic result. By considering n equally–spaced subintervals, we obtained a formula for an approximation of the definite integral that involved our variable n. As n grows large – without bound – the error shrinks to zero and we obtain the exact area.

In addition to the left, right and middle Riemann sums, also **upper and lower Riemann sums** (*boven en onder Riemann som*) can be defined. For that purpose, we consider a partition as before, and note that f has both a minimum and maximum on $[x_i, x_{i+1}]$, so there are numbers l_i and u_i in $[x_i, x_{i+1}]$ such that

$$f(l_i) \le f(x) \le f(u_i)$$

for all x in $[x_i, x_{i+1}]$. If $f(x) \ge 0$, $f(l_i)\Delta x_i$ and $f(u_i)\Delta x_i$ represent the areas of rectangles having the interval $[x_i, x_{i+1}]$ as basis and having tops passing through the lowest and highest points on the graph of f on that interval (Figure 12.8). Clearly, if A_i is the area under the graph of f and above the horizontal axis, enclosed between the straight lines $x = x_i$ and $x + x_{i+1}$, then it holds that

$$f(l_i)\Delta x_i \le A_i \le f(u_i)\Delta x_i.$$

If f is not restricted to the positive half plan, then either one or both $f(l_i)\Delta x_i$ and $f(u_i)\Delta x_i$ can be negative and will then represent the area of a rectangle lying below the x-axis. Anyhow, it will always hold that $f(l_i)\Delta x_i \le f(u_i)\Delta x_i$.

With this notation in place we can define the lower Riemann sum as

$$S_l(n) = \sum_{i=1}^{n} f(l_i)\Delta x\,,$$

and the upper Riemann sum as

$$S_u(n) = \sum_{i=1}^{n} f(u_i)\Delta x\,.$$

To illustrate the subtle difference between the left and lower Riemann sums, on the one hand, and the lower Riemann sum, for instance, on the other hand, consider Figure 12.9, where the area under the sine curve between $x = 0$ and $x = \pi$ is approximated using the latter. From this figure, it should be

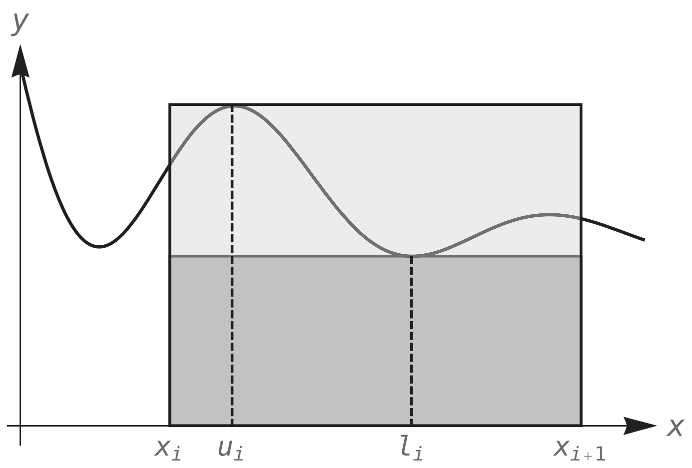


Figure 12.8: f has both a minimum and maximum on $[x_i, x_{i+1}]$.

clear that lower Riemann sum agrees with the left Riemann sum where the sine is increasing, whereas it corresponds with the right Riemann sum on the interval where the sine curve is decreasing.

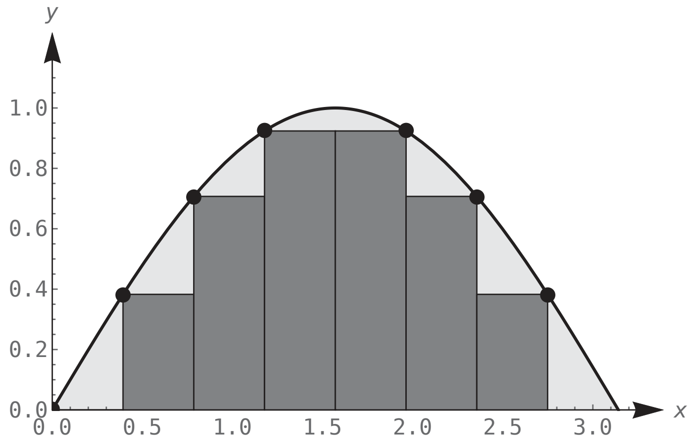


Figure 12.9: Distinction between the left and right Riemann sums and the lower Riemann sum.

12.2.3 Limits of Riemann sums

We have used limits to evaluate given definite integrals. Will this always work? We will show, given not–very–restrictive conditions, that yes, it will always work.

The previous example has shown us how we can think of a summation as a function of n. More precisely, given a definite integral $\int_a^b f(x)\,dx$, we let:

- $S_L(n) = \sum_{i=1}^{n} f(x_i)\Delta x$, be the left Riemann sum,
- $S_R(n) = \sum_{i=1}^{n} f(x_{i+1})\Delta x$, be the right Riemann sum,
- $S_M(n) = \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right)\Delta x$, be the sum of equally spaced rectangles formed using the midpoint rule,

and likewise for the lower and upper Riemann sums. Now, recall that the definition of the limit $\lim_{n\to+\infty} S_L(n) = K$ implies that given any $\epsilon > 0$, there exists $N > 0$ such that

$$\left|S_L(n) - K\right| < \epsilon,$$

when $n \geq N$.

The following theorem states that we can use any of our three rules to find the exact value of a definite integral.

Theorem 12.4 (Definite integrals and the limit of Riemann sums)
Let f be continuous on the closed interval $[a,b]$ and let $S_L(n)$, $S_R(n)$, $S_M(n)$, Δx, Δx_i and c_i be defined as before. Then:

1. $\lim_{n\to+\infty} S_L(n) = \lim_{n\to+\infty} S_R(n) = \lim_{n\to+\infty} S_M(n) = \lim_{n\to+\infty} \sum_{i=1}^{n} f(c_i)\Delta x,$

2. $\lim_{n\to+\infty}\sum_{i=1}^{n} f(c_i)\Delta x = \int_a^b f(x)\ dx$, *and*

3. $\lim_{\mathcal{L}\to 0}\sum_{i=1}^{n} f(c_i)\Delta x_i = \int_a^b f(x)\ dx$.

This theorem also goes two steps further. It states that the height of each rectangle does not have to be determined following a specific rule, but could be $f(c_i)$, where c_i is any point in the i^{th} subinterval. Furthermore, it goes on to state that the rectangles do not need to be of the same width.

Let $\mathcal{L}$ represent the length of the largest subinterval in the partition: that is, $\mathcal{L}$ is the largest of all the Δx_i's. If $\mathcal{L}$ is small, then $[a,b]$ must be partitioned into many subintervals, since all subintervals must have small lengths. Taking the limit as $\mathcal{L}$ goes to zero implies that the number n of subintervals in the partition is growing to infinity, as the largest subinterval length is becoming arbitrarily small. We then interpret the expression

$$\lim_{\mathcal{L}\to 0}\sum_{i=1}^{n} f(c_i)\Delta x_i$$

as the limit of the sum of the areas of rectangles, where the width of each rectangle can be different but getting small, and the height of each rectangle is not necessarily determined by a particular rule. The theorem states that this Riemann sum also gives the value of the definite integral of f over $[a,b]$.

Having a better understanding of the definite integral in terms of Riemann sums, we are now ready to prove the properties listed in Theorem 12.3.

We prove the fourth statement in this theorem, namely that

$$\int_a^b \big(f(x) \pm g(x)\big)\ dx = \int_a^b f(x)\ dx \pm \int_a^b g(x)\ dx\,.$$

The proofs of the other properties proceed in a similar way.

First we will prove the sum rule. From the definition of the definite integral we have,

$$\begin{aligned}
\int_a^b \big(f(x)+g(x)\big)\ dx &= \lim_{n\to+\infty}\sum_{i=1}^{n}\Big(f\big(x_i^*\big)+g\big(x_i^*\big)\Big)\Delta x \\
&= \lim_{n\to+\infty}\left(\sum_{i=1}^{n} f\big(x_i^*\big)\Delta x + \sum_{i=1}^{n} g\big(x_i^*\big)\Delta x\right) \\
&= \lim_{n\to+\infty}\sum_{i=1}^{n} f\big(x_i^*\big)\Delta x + \lim_{n\to+\infty}\sum_{i=1}^{n} g\big(x_i^*\big)\Delta x \\
&= \int_a^b f(x)\ dx + \int_a^b g(x)\ dx
\end{aligned}$$

To prove the difference formula we can either redo the above work with a minus sign instead of a plus sign or we can use the fact that we now know this is true with a plus and using the properties proved above as follows.

$$\int_a^b \big(f(x)-g(x)\big)\ dx = \int_a^b f(x) + \big(-g(x)\big)\ dx$$

$$= \int_a^b f(x)\,dx + \int_a^b \left(-g(x)\right)\,dx$$

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

By resorting to Riemann sums we can also prove some properties related to the magnitude of a definite integral. These are listed in the following theorem.

Theorem 12.5 (Properties of the magnitude of a definite integral)
Let f and g be defined on a closed interval I that contains the values a and b, and let m and M be constants. The following hold:

1. *If* $f(x) \geq 0$ *for* $a \leq x \leq b$*, then*
$$\int_a^b f(x)\,dx \geq 0.$$
2. *If* $f(x) \geq g(x)$ *for* $a \leq x \leq b$ *then*
$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx.$$
3. *If* $m \leq f(x) \leq M$ *for* $a \leq x \leq b$ *then*
$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a).$$
4. $$\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.$$

We will prove the first property in this theorem. From the definition of the definite integral we have

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

where $\Delta x = \frac{b-a}{n}$. Now, by assumption $f(x) \geq 0$ and we also have $\Delta x > 0$ and so we know that

$$\sum_{i=1}^n f\left(x_i^*\right)\Delta x \geq 0\,.$$

So, from the basic properties of limits we then have

$$\lim_{n\to+\infty} \sum_{i=1}^n f\left(x_i^*\right)\Delta x \geq \lim_{n\to+\infty} 0 = 0\,.$$

But the left side is exactly the definition of the integral and so we have

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

We now know of a way to evaluate a definite integral using limits; in the next section we will see how the fundamental theorem of calculus makes the process simpler. The key feature of this theorem is its connection between the indefinite integral and the definite integral.

Lebesgue integration

The integral we study within the framework of this course, the so-called Riemann integral, is just one kind of integral that has been proposed. While the Riemann integral considers the area under a curve as made out of vertical rectangles, the Lebesgue definition considers horizontal slabs that are not necessarily just rectangles, and so it is more flexible. For this reason, the Lebesgue definition makes it possible to calculate integrals for a broader class of functions. How the Lebesgue integral differs from the Riemann integral is illustrated in Figure 12.10 for a function f. Essentially, to compute the latter, one partitions the domain of f into subintervals, while for the latter one partitions the range of f.

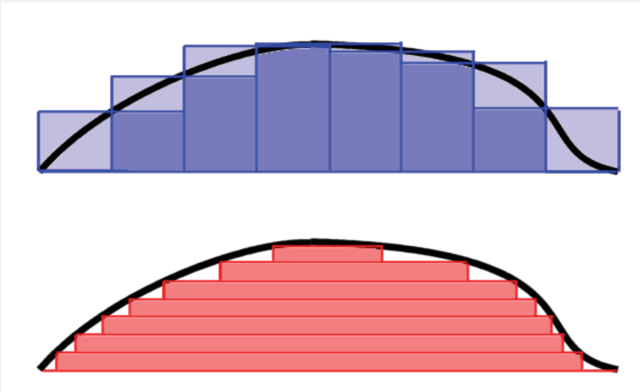


Figure 12.10: Riemann (top) versus Lebesgue integration (bottom).

12.3 The fundamental theorem of calculus

12.3.1 Mean value theorem for definite integrals

Consider the graph of a function f in Figure 12.11(a) and the area defined by $\int_1^4 f(x)\,dx$. In Figure 12.11(b), the height of the rectangle is greater than f on $[1,4]$, hence the area of this rectangle is is greater than $\int_0^4 f(x)\,dx$. In Figure 12.11(c), the height of the rectangle is smaller than f on $[1,4]$, hence the area of this rectangle is less than $\int_1^4 f(x)\,dx$. Finally, in Figure 12.11(d) the height of the rectangle is such that the area of the rectangle is exactly that of $\int_0^4 f(x)\,dx$. Since rectangles that are too big, as in Figure 12.11(b), and rectangles that are too little, as in Figure 12.11(c), give areas greater/lesser than $\int_1^4 f(x)\,dx$, it makes sense that there is a rectangle, whose top intersects $f(x)$ somewhere on $[1,4]$, whose area is exactly that of the definite integral.

We state this idea formally in a theorem.

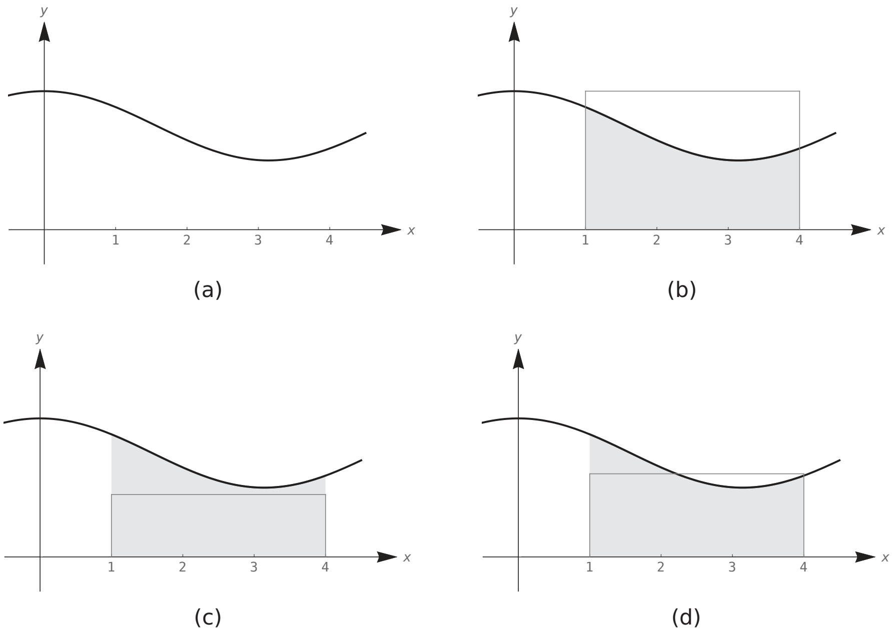


Figure 12.11: The graph of a function f (a) and differently sized rectangles give upper and lower bounds on $\int_1^4 f(x)\,dx$ (b-c).

Theorem 12.6 (The mean value theorem of integration)
Let f be continuous on $[a,b]$. There exists a value c in $[a,b]$ such that

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

This is an existential statement; c exists, but we do not provide a method of finding it. Theorem 12.6 is directly connected to the mean value theorem of differentiation (Theorem 10.4).

Let us prove this theorem by considering a more general formulation. Namely, if $f : [a,b] \to \mathbb{R}$ is continuous and g is an integrable function that does not change sign on $[a,b]$, then there exists c in $]\,a,b\,[$ such that

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx. \tag{12.5}$$

Clearly, we obtain the expression used in Theorem 12.6 by letting $g(x) = 1$.

Now to prove Equation (12.5). Let us assume that $f : [a,b] \to \mathbb{R}$ is continuous and g is a nonnegative integrable function on $[a,b]$. By the extreme value theorem (Theorem 10.1), there exists m and M such that for each x in $[a,b]$, it holds that $m \le f(x) \le M$ and $f[a,b] = [m,M]$. Since g is nonnegative, we may write that

$$m\int_a^b g(x)\,dx \leqslant \int_a^b f(x)g(x)\,dx \leqslant M\int_a^b g(x)\,dx.$$

Now let

$$I = \int_a^b g(x)\,dx.$$

Obviously, if $I = 0$, we are done since

$$0 \leqslant \int_a^b f(x)g(x)\,dx \leqslant 0$$

means

$$\int_a^b f(x)g(x)\,dx = 0,$$

so for any c in $]\,a, b\,[$,

$$\int_a^b f(x)g(x)\,dx = f(c)I = 0.$$

If $I \neq 0$, then

$$m \leqslant \frac{1}{I}\int_a^b f(x)g(x)\,dx \leqslant M.$$

By the intermediate value theorem, f attains every value of the interval $[m, M]$, so for some c in $]\,a, b\,[$ we have

$$f(c) = \frac{1}{I}\int_a^b f(x)g(x)\,dx,$$

that is,

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

Finally, if g is negative on $[a, b]$, then

$$M\int_a^b g(x)\,dx \leqslant \int_a^b f(x)g(x)\,dx \leqslant m\int_a^b g(x)\,dx,$$

and we still get the same result as above.

12.3.2 Main theorems

Let $f(t)$ be a continuous function defined on $[a, b]$. The definite integral $\int_a^b f(x)\,dx$ is the area under f on $[a, b]$. We can turn this concept into a function by letting the upper (or lower) bound vary.

Let $F(x) = \int_a^x f(t)\,dt$. It computes the area under f on $[a, x]$ as illustrated in Figure 12.12. We can study this function using our knowledge of the definite integral.

We can also apply calculus ideas to $F(x)$; in particular, we can compute its derivative. While this may seem like an innocuous thing to do, it has far–reaching implications, as demonstrated by the fact that the result is given as an important theorem.

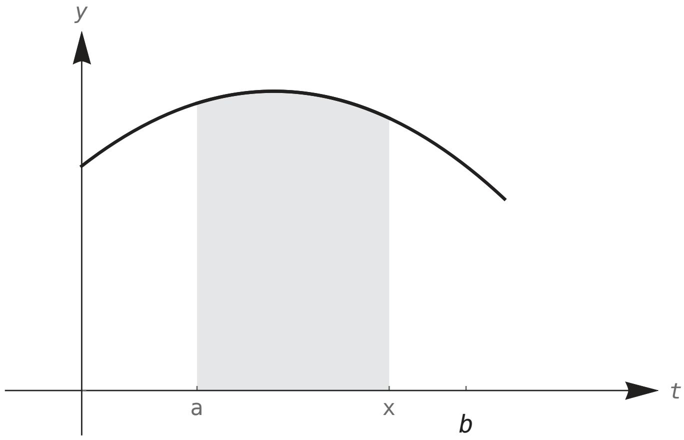


Figure 12.12: The area of the shaded region is $F(x) = \int_a^x f(t)\, dt$.

Theorem 12.7 (The fundamental theorem of calculus, Part 1)
Let f be continuous on $[a,b]$ and let $F(x) = \int_a^x f(t)\, dt$. Then F is a differentiable function on $]a,b[$, and

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

For a given $f(t)$, let us define the function $F(x)$ as

$$F(x) = \int_a^x f(t)\, dt.$$

For any two numbers x_1 and $x_1 + \Delta x$ in $[a,b]$, we have

$$F(x_1) = \int_a^{x_1} f(t)\, dt$$

and likewise

$$F(x_1 + \Delta x) = \int_a^{x_1+\Delta x} f(t)\, dt.$$

Subtracting these two equalities yields

$$F(x_1 + \Delta x) - F(x_1) = \int_a^{x_1+\Delta x} f(t)\, dt - \int_a^{x_1} f(t)\, dt. \tag{12.6}$$

Using Theorem 12.3 we can rewrite the right hand side of Equation (12.6) as

$$\int_a^{x_1+\Delta x} f(t)\, dt - \int_a^{x_1} f(t)\, dt = \int_a^{x_1+\Delta x} f(t)\, dt + \int_{x_1}^{a} f(t)\, dt = \int_{x_1}^{x_1+\Delta x} f(t)\, dt.$$

Hence, Equation (12.6) becomes

$$F(x_1 + \Delta x) - F(x_1) = \int_{x_1}^{x_1+\Delta x} f(t)\,dt. \tag{12.7}$$

According to the mean value theorem for integration (Theorem 12.6), there exists a real number $c \in [x_1, x_1 + \Delta x]$ such that

$$\int_{x_1}^{x_1+\Delta x} f(t)\,dt = f(c) \cdot \Delta x.$$

This expression allows us to rewrite Equation (12.7) as

$$F(x_1 + \Delta x) - F(x_1) = f(c) \cdot \Delta x. \tag{12.8}$$

Note that we just write c in order not to overload the notation, but one should keep in mind that, for a given function f, the value of c depends on x_1 and on Δx, though it is always confined to the interval $[x_1, x_1 + \Delta x]$.

Now, dividing both sides of Equation (12.8) by Δx gives

$$\frac{F(x_1 + \Delta x) - F(x_1)}{\Delta x} = f(c),$$

whose left side is the difference quotient for F at x_1. So, let us take the limit as $\Delta x \to 0$ on both sides of the equation. This yields:

$$\lim_{\Delta x \to 0} \frac{F(x_1 + \Delta x) - F(x_1)}{\Delta x} = \lim_{\Delta x \to 0} f(c). \tag{12.9}$$

Clearly, the expression on the left side of the resulting equation is the definition of the derivative of F at x_1, so we may rewrite Equation (12.9) as

$$F'(x_1) = \lim_{\Delta x \to 0} f(c). \tag{12.10}$$

To find the limit on the right side of Equation (12.10), we resort to the squeeze theorem (Theorem 8.5). The number c is in the interval $[x_1, x_1 + \Delta x]$, so $x_1 \leq c \leq x_1 + \Delta x$. Besides, it holds that

$$\lim_{\Delta x \to 0} x_1 = x_1$$

and

$$\lim_{\Delta x \to 0} (x_1 + \Delta x) = x_1.$$

Therefore, according to the squeeze theorem, it must hold that

$$\lim_{\Delta x \to 0} c = x_1.$$

Consequently, we may rewrite Equation (12.10) as

$$F'(x_1) = \lim_{c \to x_1} f(c).$$

The function f is continuous at c, so the limit can be taken inside the function. In this way, we get

$$F'(x_1) = f(x_1),$$

which completes the proof.

To illustrate this theorem, let us consider

$$F(x) = \int_{-5}^{x} (t^2 + \sin(t))\ dt$$

and try to find $F'(x)$.

Using Theorem 12.7, we immediately have $F'(x) = x^2 + \sin(x)$. This simple example reveals that $F(x)$ is an antiderivative of $x^2 + \sin(x)$! Therefore, $F(x) = x^3/3 - \cos(x) + C$ for some value of C. We have done more, however, than found a complicated way of computing an antiderivative. Consider a function f defined on an open interval containing a, b and c. Suppose we want to compute $\int_a^b f(t)\ dt$. First, let

$$F(x) = \int_c^x f(t)\ dt.$$

Using the properties of the definite integral (Theorem 12.3), we know

$$\begin{aligned}
\int_a^b f(t)\ dt &= \int_a^c f(t)\ dt + \int_c^b f(t)\ dt \\
&= -\int_c^a f(t)\ dt + \int_c^b f(t)\ dt \\
&= -F(a) + F(b) \\
&= F(b) - F(a).
\end{aligned}$$

We now see how indefinite integrals and definite integrals are related: we can evaluate a definite integral using antiderivatives. This is the second part of the fundamental theorem of calculus.

Theorem 12.8 (The fundamental theorem of calculus, Part 2)
Let f be continuous on $[a,b]$ and let F be any antiderivative of f. Then

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

We will rely on Riemann sums to prove this theorem in a more rigorous way. For that purpose, let f be integrable on the interval $[a,b]$, and let f admit an antiderivative F on $[a,b]$. Consider the quantity $F(b) - F(a)$ and let there be a partition of size $\mathcal{L}$ with numbers $x_1, x_2, \ldots, x_{n+1}$ such that

$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b.$$

Clearly,

$$F(b) - F(a) = F(x_{n+1}) - F(x_1).$$

Now, for $i = 2, \ldots, n$ we add each $F(x_i)$ along with its additive inverse, so that the resulting quantity is equal:

$$\begin{aligned}
F(b) - F(a) &= F(x_{n+1}) + [-F(x_n) + F(x_n)] + \cdots + [-F(x_2) + F(x_2)] - F(x_1) \\
&= [F(x_{n+1}) - F(x_n)] + [F(x_n) - F(x_{n-1})] + \cdots + [F(x_3) - F(x_2)] + [F(x_2) - F(x_1)].
\end{aligned}$$

Or shorter:

$$F(b)-F(a)=\sum_{i=1}^{n}[F(x_{i+1})-F(x_i)]. \tag{12.11}$$

Inspecting the right hand side of this equation reminds us to the mean value theorem of differentiation (Theorem 10.4). Indeed, it tells us that

$$F'(c)=\frac{F(b)-F(a)}{b-a},$$

where $c \in [a,b]$, or equivalently

$$F(b)-F(a)=F'(c)(b-a).$$

Hence, since the function F is differentiable on the interval $[a,b]$ and hence differentiable and continuous on each interval $[x_{i-1},x_i]$, we can rewrite the terms appearing in right hand side of Equation (12.11) as

$$F(x_{i+1})-F(x_i)=F'(c_i)(x_{i+1}-x_i),$$

where $c_i \in [x_{i+1},x_i]$. This allows us to rewrite Equation (12.11) as

$$F(b)-F(a)=\sum_{i=1}^{n}[F'(c_i)(x_{i+1}-x_i)].$$

Moreover, our assumption that F is an antiderivative of f implies that $F'(c_i)=f(c_i)$. Hence, letting $x_{i+1}-x_i=\Delta x_i$, we get

$$F(b)-F(a)=\sum_{i=1}^{n}[f(c_i)(\Delta x_i)]. \tag{12.12}$$

Essentially, we are describing with this expression the area of a rectangle, with the width times the height, and we are adding the areas together. By taking the limit of the expression as the norm of the partitions approaches zero, we arrive at the Riemann integral. We know that this limit exists because f was assumed to be integrable. So, we take the limit on both sides of Equation (12.12), to obtain

$$\lim_{\mathcal{L}\to 0}\big(F(b)-F(a)\big)=\lim_{\mathcal{L}\to 0}\sum_{i=1}^{n}[f(c_i)(\Delta x_i)].$$

Neither $F(b)$ nor $F(a)$ is dependent on $\mathcal{L}$, so the limit on the left side remains $F(b)-F(a)$. Furthermore, the expression on the right side of the equation defines the integral over f from a to b (Theorem 12.4). Therefore, we obtain

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

which completes the proof.

Example 12.8

We spent a great deal of time in the previous section studying $\int_0^4 (4x-x^2)\,dx$. Using the fundamental theorem of calculus, evaluate this definite integral.

Solution

We need an antiderivative of $f(x)=4x-x^2$. All antiderivatives of f have the form

$$F(x)=2x^2-\frac{1}{3}x^3+C;$$

for simplicity, choose $C = 0$.

The fundamental theorem of calculus states

$$\int_0^4 (4x - x^2)\, dx = F(4) - F(0) = \left(2(4)^2 - \frac{1}{3}4^3 - (0-0)\right) = 32 - \frac{64}{3} = \frac{32}{3}.$$

This is the same answer we obtained using limits in the previous section, just with much less work.

A special notation is often used in the process of evaluating definite integrals using the fundamental theorem of calculus. Instead of explicitly writing $F(b) - F(a)$, the notation $F(x)\Big|_a^b$ is used. Also note that any antiderivative $F(x)$ can be chosen when using the fundamental theorem of calculus to evaluate a definite integral, meaning any value of C can be picked. The constant always cancels out of the expression when evaluating $F(b) - F(a)$, so it does not matter what value is picked. This being the case, we might as well let $C = 0$.

Example 12.9

Evaluate the following definite integrals.

1. $\int_{-2}^{2} x^3\, dx$ 2. $\int_0^{\pi} \sin(x)\, dx$ 3. $\int_0^5 e^t\, dt$ 4. $\int_4^9 \sqrt{u}\, du$ 5. $\int_1^5 2\, dx$

Solution

1. $\displaystyle\int_{-2}^{2} x^3\, dx = \frac{1}{4}x^4\Big|_{-2}^{2} = \left(\frac{1}{4}2^4\right) - \left(\frac{1}{4}(-2)^4\right) = 0.$

2. $\displaystyle\int_0^{\pi} \sin(x)\, dx = -\cos(x)\Big|_0^{\pi} = -\cos(\pi) - (-\cos 0) = 1 + 1 = 2.$ So, the area under one hump of a sine curve is 2.

3. $\displaystyle\int_0^5 e^t\, dt = e^t\Big|_0^5 = e^5 - e^0 = e^5 - 1 \approx 147.41.$

4. $\displaystyle\int_4^9 \sqrt{u}\, du = \int_4^9 u^{\frac{1}{2}}\, du = \frac{2}{3}u^{\frac{3}{2}}\Big|_4^9 = \frac{2}{3}\left(9^{\frac{3}{2}} - 4^{\frac{3}{2}}\right) = \frac{2}{3}(27 - 8) = \frac{38}{3}.$

5. $\displaystyle\int_1^5 2\, dx = 2x\Big|_1^5 = 2(5) - 2 = 2(5-1) = 8.$

This last integral in Example 12.9 is interesting; the integrand is a constant function, hence we are finding the area of a rectangle with width $(5-1) = 4$ and height 2. Notice how the evaluation of the definite integral led to $2(4) = 8$.

In general, if c is a constant, then

$$\int_a^b c\ dx = c(b-a).$$

12.3.3 Motion and the fundamental theorem of calculus

We established, starting in Section 9.1.4, that the derivative of a position function is a velocity function, and the derivative of a velocity function is an acceleration function. Now consider definite integrals of velocity and acceleration functions. Specifically, if $v(t)$ is a velocity function, what does $\int_a^b v(t)\ dt$ mean?

The fundamental theorem of calculus states that

$$\int_a^b v(t)\ dt = V(b) - V(a),$$

where $V(t)$ is any antiderivative of $v(t)$. Since $v(t)$ is a velocity function, $V(t)$ must be a position function, and $V(b)-V(a)$ measures a change in position, or **displacement** (*verplaatsing*).

Example 12.10

A ball is thrown straight up with velocity given by $v(t) = -32t + 20$m/s, where t is measured in seconds. Find, and interpret, $\int_0^1 v(t)\ dt$.

Solution

Using the fundamental theorem of calculus, we have

$$\begin{aligned}\int_0^1 v(t)\ dt &= \int_0^1 (-32t+20)\ dt \\ &= -16t^2 + 20t\Big|_0^1 \\ &= 4.\end{aligned}$$

Thus if a ball is thrown straight up into the air with velocity $v(t) = -32t + 20$, the height of the ball, 1 second later, will be 4 metres above the initial height.

Integrating a rate of change function gives total change. Velocity is the rate of position change; integrating velocity gives the total change of position, i.e., displacement.

Integrating a speed function gives a similar, though different, result. Speed is also the rate of position change, but does not account for direction. So integrating a speed function gives total change of position, without the possibility of negative position change. Hence the integral of a speed function gives **distance travelled** (*afgelegde afstand*).

12.3.4 The fundamental theorem of calculus and the chain rule

Using other notation, we may write Part 1 of the fundamental theorem of calculus as

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

While we have just practised evaluating definite integrals, sometimes finding antiderivatives is impossible and we need to rely on other techniques to approximate the value of a definite integral. Functions written as

$$F(x) = \int_a^x f(t)\,dt$$

are useful in such situations.

It may be of further use to compose such a function with another. As an example, we may compose $F(x)$ with $g(x)$ to get

$$F\big(g(x)\big) = \int_a^{g(x)} f(t)\,dt.$$

What is the derivative of such a function? The chain rule can be employed to find

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'\big(g(x)\big)g'(x) = f\big(g(x)\big)g'(x).$$

An example will help us understand this.

Example 12.11

Find the derivative of

1. $F(x) = \int_2^{x^2} \ln(t)\,dt$
2. $F(x) = \int_{\cos(x)}^{5} t^3\,dt.$

Solution

1. We can view $F(x)$ as being the function $G(x) = \int_2^x \ln(t)\,dt$ composed with $h(x) = x^2$; that is, $F(x) = G\big(h(x)\big)$. The fundamental theorem of calculus states that $G'(x) = \ln(x)$. The chain rule gives us

$$\begin{aligned} F'(x) &= G'\big(h(x)\big)h'(x) \\ &= \ln(h(x))h'(x) \\ &= \ln(x^2)2x \\ &= 2x\ln(x^2). \end{aligned}$$

   Normally, of course, the steps defining $G(x)$ and $h(x)$ are skipped.

2. Note that $F(x) = -\int_5^{\cos(x)} t^3\,dt$. Viewed this way, the derivative of F is straightforward:

$$F'(x) = \sin(x)\cos^3(x).$$

12.3.5 Average value

Recognize that the mean value theorem can be rewritten as

$$f(c) = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

for some value of c in $[a,b]$. Next, partition the interval $[a,b]$ into n equally spaced subintervals, $a = x_1 < x_2 < \ldots < x_{n+1} = b$ and choose any c_i in $[x_i, x_{i+1}]$. The average of the numbers $f(c_1)$, $f(c_2)$, $\ldots$, $f(c_n)$ is:

$$\frac{1}{n}\Big(f(c_1)+f(c_2)+\ldots+f(c_n)\Big) = \frac{1}{n}\sum_{i=1}^{n} f(c_i).$$

Multiply this last expression by 1 in the form of $\frac{(b-a)}{(b-a)}$:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} f(c_i) &= \sum_{i=1}^{n} f(c_i)\frac{1}{n}\frac{(b-a)}{(b-a)} \\
&= \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\frac{b-a}{n} \\
&= \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x
\end{aligned}$$

where $\Delta x = (b-a)/n$. Now take the limit as $n \to +\infty$:

$$\lim_{n\to+\infty} \frac{1}{b-a}\sum_{i=1}^{n} f(c_i)\Delta x = \frac{1}{b-a}\int_a^b f(x)\,dx = f(c).$$

This tells us this: when we evaluate f at n (somewhat) equally spaced points in $[a,b]$, the average value of these samples is $f(c)$ as $n \to +\infty$.

This leads us to a definition.

Definition 12.5 (The average value of *f* on *[a,b]*)
Let f be continuous on $[a,b]$. The **average value of** f (*gemiddelde functiewaarde*) on $[a,b]$ is $f(c)$, where c is a value in $[a,b]$ guaranteed by the mean value theorem. I.e.,

$$\text{Average Value of } f \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

An application of this definition is given in the following example.

Example 12.12

An object moves back and forth along a straight line with a velocity given by $v(t) = (t-1)^2$ on $[0,3]$, where t is measured in seconds and $v(t)$ is measured in m/s.

1. What is the average velocity of the object?
2. What was the displacement of the object?

Solution

1. By Definition 12.5, the average velocity is:

$$\frac{1}{3-0}\int_0^3 (t-1)^2\ dt = \frac{1}{3}\int_0^3 \left(t^2-2t+1\right)\ dt = \frac{1}{3}\left(\frac{1}{3}t^3-t^2+t\right)\bigg|_0^3 = 1\text{ m/s.}$$

2. We calculate this by integrating its velocity function: $\int_0^3 (t-1)^2\ dt = 3$ m. Its final position was 3 meter from its initial position after 3 seconds: its average velocity was 1 m/s.

This section has laid the groundwork for a lot of great mathematics to follow. The most important lesson is this: definite integrals can be evaluated using antiderivatives. Since the previous section established that definite integrals are the limit of Riemann sums, we can later create Riemann sums to approximate values other than area under the curve, convert the sums to definite integrals, then evaluate these using the fundamental theorem of calculus. This will allow us to compute the work done by a variable force, the volume of certain solids, the arc length of curves, and more.

The downside is this: generally speaking, computing antiderivatives is much more difficult than computing derivatives. For that reason, we will see in Section 12.6.2 how to approximate the value of definite integrals beyond using the left hand, right hand and midpoint rules. These techniques are invaluable when antiderivatives cannot be computed, or when the actual function f is unknown and all we know is the value of f at certain x-values. But first, we will study techniques of finding antiderivatives analytically so that a wide variety of definite integrals can be evaluated.

12.4 Techniques of antidifferentiation

This chapter is devoted to exploring techniques of antidifferentiation. While not every function has an antiderivative in terms of elementary functions like polynomial, exponential or trigonometric functions, we can still find antiderivatives of a wide variety of functions.

12.4.1 Substitution

12.4.1.1 Rationale

Essentially, integration by **substitution** (*substitutie*) allows us to undo the chain rule. Its underlying principle is to rewrite a complicated integral of the form $\int f(x)\ dx$ as a not–so–complicated integral $\int h(u)\ du$.

For instance, consider

$$\int (20x+30)(x^2+3x-5)^9\ dx.$$

Arguably the most complicated part of the integrand is $(x^2+3x-5)^9$. We wish to make this simpler; we do so through a substitution. Let $u = x^2+3x-5$. Thus

$$(x^2+3x-5)^9 = u^9.$$

We have established u as a function of x, so now consider the differential of u:

$$du = (2x+3)dx.$$

Let us return now to the original integral and do some substitutions through algebra:

$$\begin{aligned}\int (20x+30)(x^2+3x-5)^9\ dx &= \int 10(2x+3)(x^2+3x-5)^9\ dx\\ &= \int 10(\underbrace{x^2+3x-5}_{u})^9\ \underbrace{(2x+3)\ dx}_{du}\\ &= \int 10u^9\ du\\ &= u^{10}+C \qquad \text{(Replace } u \text{ with } x^2+3x-5.)\\ &= (x^2+3x-5)^{10}+C\end{aligned}$$

In general, let $F(x)$ and $g(x)$ be differentiable functions and consider the derivative of their composition:

$$\frac{d}{dx}\Big(F\big(g(x)\big)\Big) = F'(g(x))g'(x).$$

Thus

$$\int F'(g(x))g'(x)\ dx = F(g(x))+C.$$

Integration by substitution works by recognizing the inside function $g(x)$ and replacing it with a variable. By setting $u = g(x)$, we can rewrite the derivative as

$$\frac{d}{dx}\Big(F\big(u\big)\Big) = F'(u)u'.$$

Since $du = g'(x)dx$, we can rewrite the above integral as

$$\int F'(g(x))g'(x)\ dx = \int F'(u)du = F(u)+C = F(g(x))+C.$$

The point of substitution is to make the integration step easy. Indeed, the step $\int F'(u)\ du = F(u)+C$ looks easy, as the antiderivative of the derivative of F is just F, plus a constant. The work involved is making the proper substitution. There is not a step–by–step process that one can memorize; rather, experience will be one's guide. To gain experience, we now embark on some examples.

Example 12.13

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{7}{-3x+1}\ dx$, 2. $\displaystyle\int x\sin(x^2+5)\ dx$, 3. $\displaystyle\int x\sqrt{x+3}\ dx$.

Solution

1. View the integrand as the composition of functions $f(g(x))$, where $f(x) = 7/x$ and $g(x) = -3x+1$. Then, we let $u = -3x+1$, the inside function. Thus $du = -3dx$. The integrand lacks a -3; hence divide the previous equation by -3 to obtain $-du/3 = dx$. We can now evaluate the integral through substitution.

$$\int \frac{7}{-3x+1}\ dx = \int \frac{7}{u}\frac{du}{(-3)}$$

$$
\begin{aligned}
&= \frac{-7}{3}\int \frac{du}{u} \\
&= \frac{-7}{3}\ln|u| + C \\
&= -\frac{7}{3}\ln|-3x+1| + C.
\end{aligned}
$$

2. We choose to let u be the inside function of $\sin(x^2+5)$. So, let $u = x^2+5$, hence $du = 2x\,dx$. The integrand has an $x\,dx$ term, but not a $2x\,dx$ term. We can divide both sides of the du expression by 2:

$$du = 2x\,dx \quad \Rightarrow \quad \frac{1}{2}du = x\,dx.$$

   We can now substitute.

$$
\begin{aligned}
\int x\sin(x^2+5)\ dx &= \int \sin(\underbrace{x^2+5}_{u})\underbrace{x\ dx}_{\frac{1}{2}du} \\
&= \int \frac{1}{2}\sin(u)\ du \\
&= -\frac{1}{2}\cos(u) + C \qquad \text{(Now replace } u \text{ with } x^2+5.\text{)} \\
&= -\frac{1}{2}\cos(x^2+5) + C.
\end{aligned}
$$

   Thus

$$\int x\sin(x^2+5)\ dx = -\frac{1}{2}\cos(x^2+5) + C.$$

3. Recognizing the composition of functions, set $u = x+3$. Then $du = dx$, giving what seems initially to be a simple substitution. But at this stage, we have:

$$\int x\sqrt{x+3}\ dx = \int x\sqrt{u}\ du.$$

   We cannot evaluate an integral that has both an x and an u in it. We need to convert the x to an expression involving just u.

   Since we set $u = x+3$, we can also state that $u-3 = x$. Thus we can replace x in the integrand with $u-3$. It will also be helpful to rewrite $\sqrt{u}$ as $u^{\frac{1}{2}}$.

$$
\begin{aligned}
\int x\sqrt{x+3}\ dx &= \int (u-3)u^{\frac{1}{2}}\ du \\
&= \int \left(u^{\frac{3}{2}} - 3u^{\frac{1}{2}}\right) du \\
&= \frac{2}{5}u^{\frac{5}{2}} - 2u^{\frac{3}{2}} + C \\
&= \frac{2}{5}(x+3)^{\frac{5}{2}} - 2(x+3)^{\frac{3}{2}} + C
\end{aligned}
$$

12.4.1.2 Integrals involving trigonometric functions

Integration by substitution can also be used to unveil the antiderivatives of the tangent, cotangent, secant and cosecant. For instance, consider the following example concerning the former function.

Example 12.14

Evaluate

$$\int \tan(x)\,dx.$$

Solution

Rewrite $\tan(x)$ as $\sin(x)/\cos(x)$. While the presence of a composition of functions may not be immediately obvious, recognize that $\cos(x)$ is inside the $1/x$ function. Therefore, we see if setting $u=\cos(x)$ returns usable results. We have that $du=-\sin(x)\,dx$, hence $-du=\sin(x)\,dx$. We can integrate:

$$\begin{aligned}\int \tan(x)\,dx &= \int \frac{\sin(x)}{\cos(x)}\,dx \\ &= \int \underbrace{\frac{1}{\cos(x)}}_{1/u}\underbrace{\sin(x)\,dx}_{-du} \\ &= \int \frac{-1}{u}\,du \\ &= -\ln|u|+C \\ &= -\ln|\cos(x)|+C.\end{aligned}$$

This can be simplified even further by bringing the -1 inside the logarithm as a power of $\cos(x)$, as in:

$$\begin{aligned}-\ln|\cos(x)|+C &= \ln|(\cos(x))^{-1}|+C \\ &= \ln\left|\frac{1}{\cos(x)}\right|+C \\ &= \ln|\sec(x)|+C.\end{aligned}$$

Thus the result they give is $\int \tan(x)\,dx = \ln|\sec(x)|+C$.

We can use similar techniques in Example 12.14 to find antiderivatives of the other trigonometric functions. In this way, one finds:

1. $\displaystyle\int \tan(x)\,dx = -\ln|\cos(x)|+C$
2. $\displaystyle\int \cot(x)\,dx = \ln|\sin(x)|+C$
3. $\displaystyle\int \sec(x)\,dx = \ln|\sec(x)+\tan(x)|+C$
4. $\displaystyle\int \csc(x)\,dx = -\ln|\csc(x)+\cot(x)|+C$

Likewise, we can find antiderivatives of hyperbolic functions:

1. $\displaystyle\int \tanh(x)\, dx = \ln(\cosh(x)) + C$

2. $\displaystyle\int \coth(x)\, dx = \ln|\sinh(x)| + C$

Using the power-reducing formulas we have seen in Chapter 5 (Theorem 5.12), we can also tackle integrals involving powers of trigonometric and hyperbolic functions.

Example 12.15

Evaluate

$$\int \cos^2(x)\, dx.$$

Solution

We have a composition of functions as $\cos^2(x) = \big(\cos(x)\big)^2$. However, setting $u = \cos(x)$ means $du = -\sin(x)\, dx$, which we do not have in the integral. So, let us use Theorem 5.12, which states

$$\cos^2(x) = \frac{1+\cos(2x)}{2}.$$

The right hand side of this equation is not difficult to integrate. We have:

$$\begin{aligned}\int \cos^2(x)\, dx &= \int \frac{1+\cos(2x)}{2}\, dx\\ &= \int \left(\frac{1}{2} + \frac{1}{2}\cos(2x)\right) dx.\end{aligned}$$

So, we easily find:

$$\begin{aligned}&= \frac{1}{2}x + \frac{1}{2}\frac{\sin(2x)}{2} + C\\ &= \frac{1}{2}x + \frac{\sin(2x)}{4} + C.\end{aligned}$$

We will make significant use of this power–reducing technique in future sections.

12.4.1.3 Integrals leading to inverse trigonometric and hyperbolic functions

When studying derivatives of inverse functions, we learned that

$$\frac{d}{dx}\big(\arctan(x)\big) = \frac{1}{1+x^2}.$$

Applying the chain rule to this is not difficult. For instance, in general, we have

$$\frac{d}{dx}\big(\arctan(ax)\big) = \frac{a}{1+a^2x^2}.$$

This result can be used to evaluate

$$\int \frac{1}{a^2+x^2}\, dx$$

.

For that purpose, we rewrite this integral as

$$\frac{1}{a^2}\int \frac{1}{1+\left(\frac{x}{a}\right)^2}\,dx.$$

This can now be integrated using substitution. Set $u = x/a$, hence $du = dx/a$ or $dx = a\,du$. Thus

$$\begin{aligned}\int \frac{1}{a^2+x^2}\,dx &= \frac{1}{a}\int \frac{1}{1+u^2}\,du\\ &= \frac{1}{a}\arctan(u)+C\\ &= \frac{1}{a}\arctan\left(\frac{x}{a}\right)+C\end{aligned}$$

This demonstrates a general technique that can be applied to other integrands that result in inverse trigonometric functions. More specifically, for $a > 0$, we have

$$\int \frac{1}{a^2+x^2}\,dx = \frac{1}{a}\arctan\left(\frac{x}{a}\right)+C \tag{12.13}$$

$$\int \frac{1}{\sqrt{a^2-x^2}}\,dx = \arcsin\left(\frac{x}{a}\right)+C \tag{12.14}$$

Of course, given the link between trigonometric and hyperbolic functions, similar integrands result in inverse hyperbolic functions:

$$\int \frac{1}{\sqrt{x^2-a^2}}\,dx = \operatorname{arcosh}\left(\frac{x}{a}\right)+C = \ln\left|x+\sqrt{x^2-a^2}\right|+C, \qquad \text{for } 0 < a < x, \tag{12.15}$$

$$\int \frac{1}{\sqrt{x^2+a^2}}\,dx = \operatorname{arsinh}\left(\frac{x}{a}\right)+C = \ln\left|x+\sqrt{x^2+a^2}\right|+C, \qquad \text{for } a > 0, \tag{12.16}$$

$$\int \frac{1}{a^2-x^2}\,dx = \begin{cases}\frac{1}{a}\operatorname{artanh}\left(\frac{x}{a}\right)+C, & x^2 < a^2,\\ \frac{1}{a}\operatorname{arcoth}\left(\frac{x}{a}\right)+C, & a^2 < x^2\end{cases} \tag{12.17}$$

$$= \frac{1}{2a}\ln\left|\frac{a+x}{a-x}\right|+C, \tag{12.18}$$

Example 12.16

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{x^2-4x+13}\,dx$,
2. $\displaystyle\int \frac{4-x}{\sqrt{16-x^2}}\,dx$,
3. $\displaystyle\int \frac{1}{x^2-1}\,dx$,
4. $\displaystyle\int \frac{1}{\sqrt{9x^2+10}}\,dx$.

Solution

1. We start by completing the square in the denominator, i.e.

$$\frac{1}{x^2-4x+13} = \frac{1}{(x-2)^2+9}$$

We can now integrate, to arrive at

$$\int \frac{1}{x^2-4x+13}\,dx = \int \frac{1}{(x-2)^2+9}\,dx = \frac{1}{3}\arctan\left(\frac{x-2}{3}\right)+C.$$

2. This integral requires two different methods to evaluate it. We get to those methods by splitting up the integral:

$$\int \frac{4-x}{\sqrt{16-x^2}}\,dx = \int \frac{4}{\sqrt{16-x^2}}\,dx - \int \frac{x}{\sqrt{16-x^2}}\,dx.$$

The first integral is easy to handle; the second integral is handled by substitution, with $u=16-x^2$. We handle each separately.

$\displaystyle\int \frac{4}{\sqrt{16-x^2}}\,dx = 4\arcsin\left(\frac{x}{4}\right)+C.$

$\displaystyle\int \frac{x}{\sqrt{16-x^2}}\,dx$: Set $u=16-x^2$, so $du=-2xdx$ and $xdx=-du/2$. We have

$$\begin{aligned}\int \frac{x}{\sqrt{16-x^2}}\,dx &= \int \frac{-du/2}{\sqrt{u}}\\ &= -\frac{1}{2}\int \frac{1}{\sqrt{u}}\,du\\ &= -\sqrt{u}+C\\ &= -\sqrt{16-x^2}+C.\end{aligned}$$

Combining these together, we have

$$\int \frac{4-x}{\sqrt{16-x^2}}\,dx = 4\arcsin\left(\frac{x}{4}\right)+\sqrt{16-x^2}+C.$$

3. Multiplying the numerator and denominator by (-1) gives:

$$\int \frac{1}{x^2-1}\,dx = \int \frac{-1}{1-x^2}\,dx.$$

The second integral can be solved directly using Equation (12.18), with $a = 1$. Thus

$$\begin{aligned}\int \frac{1}{x^2-1}\,dx &= -\int \frac{1}{1-x^2}\,dx \\ &= \begin{cases} -\operatorname{artanh}(x)+C & x^2<1 \\ -\operatorname{arcosh}(x)+C & 1<x^2 \end{cases} \\ &= -\frac{1}{2}\ln\left|\frac{x+1}{x-1}\right|+C \\ &= \frac{1}{2}\ln\left|\frac{x-1}{x+1}\right|+C. \end{aligned} \tag{12.19}$$

4. This requires a substitution, then Equation (12.16) can be used.

   Let $u = 3x$, hence $du = 3dx$. We have

$$\int \frac{1}{\sqrt{9x^2+10}}\,dx = \frac{1}{3}\int \frac{1}{\sqrt{u^2+10}}\,du.$$

   Note $a^2 = 10$, hence $a = \sqrt{10}$. We immediately obtain

$$\int \frac{1}{\sqrt{9x^2+10}}\,dx = \frac{1}{3}\operatorname{arsinh}\left(\frac{3x}{\sqrt{10}}\right)+C = \frac{1}{3}\ln\left|3x+\sqrt{9x^2+10}\right|+C$$

12.4.1.4 Substitution and definite integration

Definite integrals that require substitution can be calculated using the following workflow:

1. Start with a definite integral $\int_a^b f(x)\,dx$ that requires substitution.
2. Ignore the bounds; use substitution to evaluate $\int f(x)\,dx$ and find an antiderivative $F(x)$.
3. Evaluate $F(x)$ at the bounds; that is, evaluate $F(x)\Big|_a^b = F(b) - F(a)$.

This workflow works fine, but substitution offers an alternative that is powerful and time saving. Since substitution converts integrals of the form $\int F'(g(x))g'(x)\,dx$ into an integral of the form $\int F'(u)\,du$ with the substitution of $u = g(x)$, we just have to appropriately change the bounds of a definite integral, i.e.

$$\int_a^b F'\big(g(x)\big)g'(x)\,dx = \int_{g(a)}^{g(b)} F'(u)\,du.$$

This indicates that once you convert to integrating with respect to u, you do not need to switch back to evaluating with respect to x.

Example 12.17

Evaluate

$$\int_0^{\pi/2} \sin(x)\cos(x)\,dx.$$

Solution

Let $u = g(x) = \cos(x)$, giving $du = -\sin(x)\ dx$ and hence $\sin(x)\ dx = -du$. The new upper bound is $g(\pi/2) = 0$; the new lower bound is $g(0) = 1$. Note how the lower bound is actually larger than the upper bound now. We have

$$\begin{aligned}\int_0^{\pi/2} \sin(x)\cos(x)\ dx &= \int_1^0 -u\ du \\ &= \int_0^1 u\ du \\ &= \frac{1}{2}u^2\Big|_0^1 = \frac{1}{2}.\end{aligned}$$

In Figure 12.13 we have graphed the two regions defined by our definite integrals. They bear no resemblance to each other, but they have the same area.

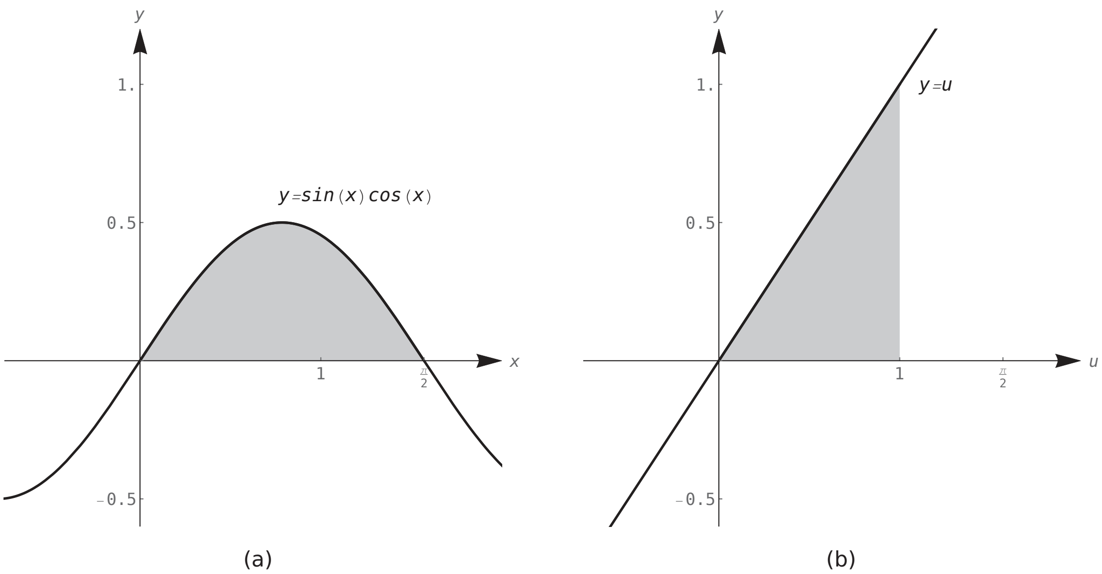


Figure 12.13: Graphing the areas defined by the definite integrals of Example 12.17.

12.4.1.5 Tangent half-angle substitution

The tangent half-angle substitution, also known as the Weierstrass substitution after Karl Weierstrass, is a substitution used for finding antiderivatives of rational functions of trigonometric functions.

For this substitution we let $t = \tan\left(\frac{x}{2}\right)$. By the double-angle formula for the sine function, we get

$$\begin{aligned}\sin(x) &= 2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right) \\ &= 2t\cos^2\left(\frac{x}{2}\right) \\ &= \frac{2t}{\sec^2\left(\frac{x}{2}\right)}\end{aligned}$$

$$= \frac{2t}{1+t^2}.$$

Similarly, by the double-angle formula for the cosine function, we easily find

$$\begin{aligned}
\cos(x) &= 1 - 2\sin^2\left(\frac{x}{2}\right) \\
&= 1 - 2t^2\cos^2\left(\frac{x}{2}\right) \\
&= 1 - \frac{2t^2}{\sec^2\left(\frac{x}{2}\right)} \\
&= 1 - \frac{2t^2}{1+t^2} \\
&= \frac{1-t^2}{1+t^2}.
\end{aligned}$$

Moreover, since

$$\begin{aligned}
\frac{dt}{dx} &= \frac{1}{2}\sec^2\left(\frac{x}{2}\right) \\
&= \frac{1+t^2}{2},
\end{aligned}$$

we get the following expression for dx:

$$dx = \frac{2}{1+t^2}\,dt\,.$$

Example 12.18

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{1+\sin(x)}\,dx$
2. $\displaystyle\int \frac{\sin(x)}{2+\cos^2(x)}\,dx$

Solution

1. Using the Weierstrass substitution, we easily find

$$\int \frac{dx}{1+\sin(x)} = \int \frac{1}{1+\left(\frac{2t}{1+t^2}\right)}\,\frac{2}{1+t^2}\,dt = \int \frac{2}{(t+1)^2}\,dt\,,$$

where the last integral can be evaluated by a change a variables. Indeed, letting $u = t+1$, we find

$$\int \frac{2}{(t+1)^2}\,dt = \frac{-2}{1+t} + C\,,$$

or in terms of the original variable x where we started from:

$$\int \frac{dx}{1+\sin(x)} = \frac{-2}{1+\tan\left(\frac{x}{2}\right)} + C\,.$$

2. Again, it is clear that the Weierstrass substitution will help us out:

$$\begin{aligned}\int \frac{\sin(x)}{2+\cos^2(x)}\,dx &= \int \frac{\left(\frac{2t}{1+t^2}\right)}{2+\left(\frac{1-t^2}{1+t^2}\,dt\right)^2}\left(\frac{2}{1+t^2}\,dt\right)\\ &= \int \frac{4t}{3t^4+2t^2+3}\,dt\,.\end{aligned}$$

It should be clear that we can recast the last integral in order to arrive an arctangent function. This can be done, by letting $v = t^2$, to get

$$\int \frac{2}{3v^2+2v+3}\,dv\,,$$

which can be rewritten, after some algebra, as

$$\frac{3}{4}\int \frac{1}{\left(\frac{3v+1}{2\sqrt{2}}\right)^2+1}\,dv\,.$$

Using the substitution $w = \frac{3v+1}{2\sqrt{2}}$, the latter integral on its turn becomes

$$\frac{\sqrt{2}}{2}\int \frac{1}{w^2+1}\,dw\,,$$

which in terms of v evaluates to

$$\frac{\sqrt{2}}{2}\arctan\left(\frac{3v+1}{2\sqrt{2}}\right).$$

Consequently, in terms of the original variable x, we arrive at

$$\int \frac{\sin(x)}{2+\cos^2(x)}\,dx = \frac{\sqrt{2}}{2}\arctan\left(\frac{3\left(\tan(\frac{x}{2})\right)^2+1}{2\sqrt{2}}\right).$$

12.4.2 Integration by parts

Here is a simple integral that we can not yet evaluate:

$$\int x\cos(x)\,dx.$$

It's a simple matter to take the derivative of the integrand using the product rule, but there is no such rule for integrals. However, this section introduces **integration by parts** (*partiële integratie*), a method of integration that is based on the product rule for derivatives. It will enable us to evaluate this integral.

The product rule says that if u and v are functions of x, then $(uv)' = u'v + uv'$. For simplicity, we have written u for $u(x)$ and v for $v(x)$. Suppose we integrate both sides with respect to x. This gives

$$\int (uv)'\,dx = \int (u'v+uv')\,dx.$$

By the fundamental theorem of calculus, the left side integrates to uv. The right side can be broken up into two integrals, and we have

$$uv = \int u'v\,dx + \int uv'\,dx.$$

Solving for the second integral we have

$$\int uv'\,dx = uv - \int u'v\,dx.$$

Using differential notation, we can write $du = u'(x)dx$ and $dv = v'(x)dx$ and the expression above can be written as follows:

$$\int u\,dv = uv - \int v\,du. \tag{12.20}$$

If our problem concerns a definite integral, we likewise arrive at

$$\int_{x=a}^{x=b} u\ dv = uv\Big|_a^b - \int_{x=a}^{x=b} v\ du.$$

Typically, we try to identify u and dv in the integral we are given, and the key is that we usually want to choose u and dv so that du is simpler than u and v is hopefully not too much more complicated than dv. This will mean that the integral on the right side of the integration by parts formula, $\int v\,du$ will be simpler to integrate than the original integral $\int u\,dv$.

Example 12.19

Evaluate the following indefinite integrals:

1. $\int x^2\cos(x)\,dx$
2. $\int e^x\cos(x)\,dx$
3. $\int \arctan(x)\,dx$
4. $\int \cos(\ln(x))\ dx$

Solution

1. Let $u = x^2$ so that $dv = \cos(x)\,dx$. Then, it follows that $du = 2x\,dx$ and $v = \sin(x)$. Equation (12.20) leads to

$$\int x^2\cos(x)\,dx = x^2\sin(x) - \int 2x\sin(x)\,dx.$$

At this point, the integral on the right is indeed simpler than the one we started with, but to evaluate it, we need to do integration by parts again. Here we choose $u = 2x$ and $dv = \sin x$, so that $du = 2dx$ and $v = -\cos(x)$. Through Equation (12.20) this yields:

$$\int x^2\cos(x)\,dx = x^2\sin(x) - \left(-2x\cos(x) - \int -2\cos(x)\,dx\right).$$

The integral all the way on the right is now something we can evaluate. It evaluates to $-2\sin x$. Then going through and simplifying, being careful to keep all the signs straight, our answer is

$$\int x^2\cos(x)\ dx = x^2\sin(x) + 2x\cos(x) - 2\sin(x) + C.$$

2. This is a classic problem. In this particular example, one can let u be either $\cos(x)$ or e^x; we choose $u = e^x$ and hence $dv = \cos(x)\,dx$. Then $du = e^x\,dx$ and $v = \sin(x)$ as shown below. Using Equation (12.20) yields

$$\int e^x \cos(x)\ dx = e^x \sin(x) - \int e^x \sin(x)\,dx.$$

The integral on the right is not much different than the one we started with, so it seems like we have gotten nowhere. Let us nonetheless keep working and apply integration by parts to the new integral, using $u = e^x$ and $dv = \sin(x)\,dx$. Then we get $du = e^x\,dx$ and $v = -\cos(x)$ and this leads us to the following:

$$\begin{aligned}\int e^x \cos(x)\,dx &= e^x \sin(x) - \left(-e^x \cos(x) - \int -e^x \cos(x)\,dx\right)\\ &= e^x \sin(x) + e^x \cos(x) - \int e^x \cos(x)\ dx.\end{aligned}$$

It seems we are back right where we started, as the right hand side contains $\int e^x \cos(x)\,dx$. But this is actually a good thing.

Add $\int e^x \cos(x)\ dx$ to both sides. This gives

$$2\int e^x \cos(x)\ dx = e^x \sin(x) + e^x \cos(x)$$

Now divide both sides by 2:

$$\int e^x \cos(x)\ dx = \frac{1}{2}\big(e^x \sin(x) + e^x \cos(x)\big).$$

Simplifying a little and adding the constant of integration, our answer is thus

$$\int e^x \cos(x)\ dx = \frac{1}{2} e^x \big(\sin(x) + \cos(x)\big) + C.$$

3. Let $u = \arctan(x)$ and $dv = dx$. Then $du = 1/(1+x^2)\,dx$ and $v = x$. Using Equation (12.20) yields

$$\int \arctan(x)\,dx = x\arctan(x) - \int \frac{x}{1+x^2}\,dx.$$

The integral on the right can be solved by substitution. Taking $u = 1+x^2$, we get $du = 2x\,dx$. The integral then becomes

$$\int \arctan(x)\,dx = x\arctan(x) - \frac{1}{2}\int \frac{1}{u}\,du.$$

The integral on the right evaluates to $\ln|u| + C$, which becomes $\ln(1+x^2) + C$, as we may drop the absolute values as $1+x^2$ is always positive. Therefore, the answer is

$$\int \arctan(x)\ dx = x\arctan(x) - \frac{1}{2}\ln(1+x^2) + C.$$

4. The integrand contains a composition of functions, leading us to think integration by parts would be beneficial. Letting $u = \cos\big(\ln(x)\big)$, we have $du = -\sin\big(\ln(x)\big)/x\ dx$, and conse-

quently $dv = dx$ and $v = x$. We then have

$$\int \cos(\ln(x))\ dx = x\cos(\ln(x)) + \int \sin(\ln(x))\ dx$$

$$= \cos(\ln(x)) + x\sin(\ln(x)) - \int \cos(\ln(x))\ dx\,.$$

So, we see that

$$\int \cos(\ln(x))\ dx = \frac{1}{2}x\big(\sin(\ln(x)) + \cos(\ln(x))\big) + C.$$

In general, integration by parts is useful for integrating certain products of functions, like $\int xe^x\,dx$ or $\int x^3\sin(x)\,dx$. It is also useful for integrals involving logarithms and inverse trigonometric functions.

As stated before, integration is generally more difficult than differentiation. We are developing tools for handling a large array of integrals, and experience will tell us when one tool is preferable/necessary over another. For instance, consider the three similar–looking integrals

$$\int xe^x\,dx, \qquad \int xe^{x^2}\,dx \qquad \text{and} \qquad \int xe^{x^3}\,dx.$$

While the first is calculated easily with integration by parts, the second is best approached with substitution. Taking things one step further, the third integral has no answer in terms of elementary functions, so none of the methods we learn in calculus will get us the exact answer.

12.4.3 Trigonometric integrals

Functions involving trigonometric functions are useful as they are good at describing periodic behavior. Here, we describe several techniques for finding antiderivatives of certain combinations of trigonometric functions.

12.4.3.1 Integrals of the form $\int \sin^m(x)\cos^n(x)\ dx$

We consider integrals of the form

$$\int \sin^m(x)\cos^n(x)\ dx,$$

where m, n are nonnegative integers. Our strategy for evaluating these integrals is to use the identity $\cos^2(x) + \sin^2(x) = 1$ to convert high powers of one trigonometric function into the other, leaving a single sine or cosine term in the integrand. This is summarized below.

1. If m is odd, then $m = 2k+1$ for some integer k. Rewrite

   $$\sin^m(x) = \sin^{2k+1}(x) = \sin^{2k}(x)\sin(x) = (\sin^2(x))^k\sin(x) = (1-\cos^2(x))^k\sin(x).$$

   Then

   $$\int \sin^m(x)\cos^n(x)\ dx = \int (1-\cos^2(x))^k\sin(x)\cos^n(x)\ dx = -\int (1-u^2)^k u^n\ du,$$

   where $u = \cos(x)$ and $du = -\sin(x)\ dx$.

2. If n is odd, then using substitutions similar to that outlined above we have

$$\int \sin^m(x)\cos^n(x)\,dx = \int u^m(1-u^2)^k\,du,$$

where $u = \sin(x)$ and $du = \cos(x)\,dx$.

3. If both m and n are even, use Theorem 5.12 to reduce the degree of the integrand. Expand the result and apply (1)-(3) again.

Let us check out how this all works in the following examples.

Example 12.20

Evaluate

$$\int \sin^5(x)\cos^9(x)\,dx.$$

Solution

The powers of both the sine and cosine terms are odd, therefore we can apply the techniques above to either power. We choose to work with the power of the cosine term.

We rewrite $\cos^9(x)$ as

$$\begin{aligned}
\cos^9(x) &= \cos^8(x)\cos(x)\\
&= (\cos^2(x))^4\cos(x)\\
&= (1-\sin^2(x))^4\cos(x)\\
&= (1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x))\cos x.
\end{aligned}$$

We rewrite the integral as

$$\int \sin^5(x)\cos^9(x)\,dx = \int \sin^5(x)\big(1-4\sin^2(x)+6\sin^4(x)-4\sin^6(x)+\sin^8(x)\big)\cos(x)\,dx.$$

Now substitute using $u = \sin(x)$ and $du = \cos(x)\,dx$ to arrive at the following integral

$$\int u^5(1-4u^2+6u^4-4u^6+u^8)\,du,$$

which can then be integrated:

$$\begin{aligned}
\int u^5(1-4u^2+6u^4-4u^6+u^8)\,du &= \int \big(u^5-4u^7+6u^9-4u^{11}+u^{13}\big)\,du\\
&= \frac{1}{6}u^6-\frac{1}{2}u^8+\frac{3}{5}u^{10}-\frac{1}{3}u^{12}+\frac{1}{14}u^{14}+C\\
&= \frac{1}{6}\sin^6(x)-\frac{1}{2}\sin^8(x)+\frac{3}{5}\sin^{10}(x)+\ldots\\
&\quad -\frac{1}{3}\sin^{12}(x)+\frac{1}{14}\sin^{14}(x)+C.
\end{aligned}$$

The work we are doing here can be a bit tedious, but the skills developed (problem solving, algebraic manipulation, etc.) are important. Nowadays problems of this sort are often solved using a computer algebra system. Mathematica, for instance, integrates $\int \sin^5(x)\cos^9(x)\,dx$ as

$$g(x) = -\frac{45\cos(2x)}{16384} - \frac{5\cos(4x)}{8192} + \frac{19\cos(6x)}{49152} + \frac{\cos(8x)}{4096} - \frac{\cos(10x)}{81920} - \frac{\cos(12x)}{24576} - \frac{\cos(14x)}{114688},$$

which clearly has a different form than our answer in Example 12.20, which we now refer to as $f(x)$. Figure 12.14 shows a graph of f and g; they are clearly not equal, but they differ only by a constant. That is $g(x) = f(x) + C$ for some constant C. So we have two different antiderivatives of the same function, meaning both answers are correct.

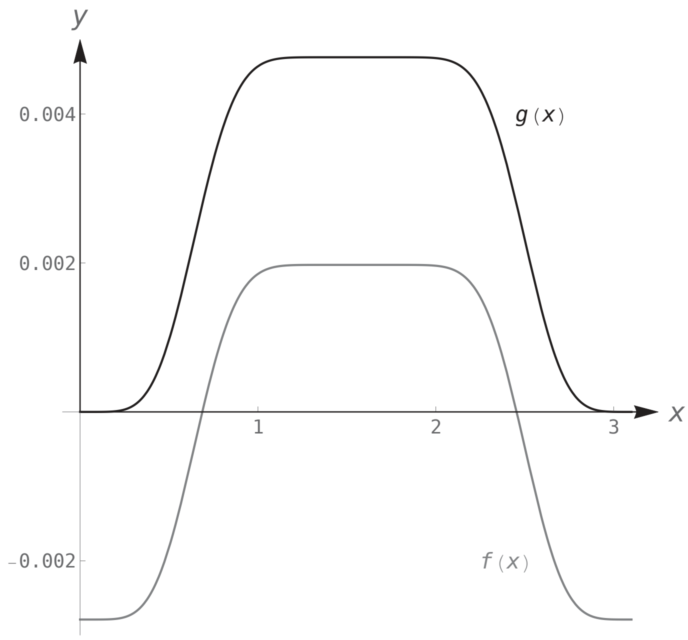


Figure 12.14: A plot of $f(x)$ and $g(x)$ from Example 12.20.

Example 12.21

Evaluate

$$\int \cos^4(x)\sin^2(x)\,dx.$$

Solution

The powers of sine and cosine are both even, so we employ the power–reducing formulas and algebra as follows.

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\,dx &= \int \left(\frac{1+\cos(2x)}{2}\right)^2\left(\frac{1-\cos(2x)}{2}\right)dx\\ &= \int \frac{1+2\cos(2x)+\cos^2(2x)}{4}\cdot\frac{1-\cos(2x)}{2}\,dx\\ &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\,dx\end{aligned}$$

The $\cos(2x)$ term is easy to integrate. The $\cos^2(2x)$ term is another trigonometric integral with an even power, requiring the power–reducing formula again. The $\cos^3(2x)$ term is a cosine function with an odd power, requiring a substitution as done before. We integrate each in turn below.

$$\int \cos(2x)\ dx = \frac{1}{2}\sin(2x) + C.$$

$$\int \cos^2(2x)\ dx = \int \frac{1+\cos(4x)}{2}\ dx = \frac{1}{2}\Big(x + \frac{1}{4}\sin(4x)\Big) + C.$$

Finally, we rewrite $\cos^3(2x)$ as

$$\cos^3(2x) = \cos^2(2x)\cos(2x) = \big(1-\sin^2(2x)\big)\cos(2x).$$

Letting $u = \sin(2x)$, we have $du = 2\cos(2x)\ dx$, hence

$$\begin{aligned}\int \cos^3(2x)\ dx &= \int \big(1-\sin^2(2x)\big)\cos(2x)\ dx \\ &= \int \frac{1}{2}(1-u^2)\ du \\ &= \frac{1}{2}\Big(u - \frac{1}{3}u^3\Big) + C \\ &= \frac{1}{2}\Big(\sin(2x) - \frac{1}{3}\sin^3(2x)\Big) + C.\end{aligned}$$

Putting all the pieces together, we have

$$\begin{aligned}\int \cos^4(x)\sin^2(x)\ dx &= \int \frac{1}{8}\big(1+\cos(2x)-\cos^2(2x)-\cos^3(2x)\big)\ dx \\ &= \frac{1}{8}\Big[x + \frac{1}{2}\sin(2x) - \frac{1}{2}\Big(x + \frac{1}{4}\sin(4x)\Big) - \frac{1}{2}\Big(\sin(2x) - \frac{1}{3}\sin^3(2x)\Big)\Big] + C \\ &= \frac{1}{8}\Big[\frac{1}{2}x - \frac{1}{8}\sin(4x) + \frac{1}{6}\sin^3(2x)\Big] + C.\end{aligned}$$

12.4.3.2 Integrals of products of sines and cosines of differing period

Integrals of the form

$$\int \sin(mx)\sin(nx)\ dx, \quad \int \cos(mx)\cos(nx)\ dx \quad \text{and} \quad \int \sin(mx)\cos(nx)\ dx$$

are best approached by first applying the product to sum formulas (Theorem 5.13).

Example 12.22

Evaluate

$$\int \sin(5x)\cos(2x)\ dx.$$

Solution

The application of the appropriate Simpson formula and subsequent integration are straightforward:

$$\int \sin(5x)\cos(2x)\,dx = \int \frac{1}{2}\Big[\sin(3x)+\sin(7x)\Big]\,dx$$
$$= -\frac{1}{6}\cos(3x) - \frac{1}{14}\cos(7x) + C$$

12.4.3.3 Integrals of the form $\int \tan^m(x)\sec^n(x)\,dx$.

When evaluating integrals of the form $\int \sin^m(x)\cos^n(x)\,dx$, the Pythagorean theorem allowed us to convert even powers of sine into even powers of cosine, and vise–versa. The same basic strategy applies to integrals of the form $\int \tan^m(x)\sec^n(x)\,dx$, albeit a bit more nuanced.

Basically, if the integrand can be manipulated to separate a $\sec^2(x)$ term with the remaining secant power even, or if a $\sec(x)\tan(x)$ term can be separated with the remaining $\tan(x)$ power even, the Pythagorean theorem can be employed, leading to a simple substitution. This strategy is outlined below.

1. If n is even, then $n = 2k$ for some integer k. Rewrite $\sec^n x$ as

   $$\sec^n(x) = \sec^{2k}(x) = \sec^{2k-2}(x)\sec^2(x) = (1+\tan^2(x))^{k-1}\sec^2(x).$$

   Then

   $$\int \tan^m(x)\sec^n(x)\,dx = \int \tan^m(x)(1+\tan^2(x))^{k-1}\sec^2(x)\,dx = \int u^m(1+u^2)^{k-1}\,du,$$

   where $u = \tan(x)$ and $du = \sec^2(x)\,dx$.

2. If m is odd, then $m = 2k+1$ for some integer k. Rewrite $\tan^m(x)\sec^n(x)$ as

   $$\begin{aligned}\tan^m(x)\sec^n(x) &= \tan^{2k+1}(x)\sec^n(x) = \tan^{2k}(x)\sec^{n-1}(x)\sec(x)\tan(x)\\ &= (\sec^2(x)-1)^k\sec^{n-1}(x)\sec(x)\tan(x).\end{aligned}$$

   Then

   $$\int \tan^m(x)\sec^n(x)\,dx = \int (\sec^2(x)-1)^k\sec^{n-1}(x)\sec(x)\tan(x)\,dx = \int (u^2-1)^k u^{n-1}\,du,$$

   where $u = \sec(x)$ and $du = \sec(x)\tan(x)\,dx$.

3. If n is odd and m is even, then $m = 2k$ for some integer k. Convert $\tan^m(x)$ to $(\sec^2(x)-1)^k$. Expand the new integrand and use Integration By Parts, with $dv = \sec^2(x)\,dx$.

4. If m is even and $n = 0$, rewrite $\tan^m(x)$ as

   $$\tan^m(x) = \tan^{m-2}(x)\tan^2(x) = \tan^{m-2}(x)(\sec^2(x)-1) = \tan^{m-2}(x)\sec^2(x) - \tan^{m-2}(x).$$

   So

   $$\int \tan^m(x)\,dx = \underbrace{\int \tan^{m-2}(x)\sec^2(x)\,dx}_{\text{apply rule \#1}} - \underbrace{\int \tan^{m-2}(x)\,dx}_{\text{apply rule \#4 again}}.$$

Example 12.23

Evaluate the following indefinite integrals:

1. $\displaystyle\int \tan^2(x)\sec^6(x)\;dx$,
2. $\displaystyle\int \tan^6(x)\;dx$.

Solution

1. Since the power of secant is even, we use rule #1 above and pull out a $\sec^2(x)$ in the integrand. We convert the remaining powers of secant into powers of tangent.

$$\begin{aligned}\int \tan^2(x)\sec^6(x)\;dx &= \int \tan^2(x)\sec^4(x)\sec^2(x)\;dx\\ &= \int \tan^2(x)\big(1+\tan^2(x)\big)^2\sec^2(x)\;dx\end{aligned}$$

Now substitute, with $u=\tan(x)$, with $du=\sec^2(x)\;dx$:

$$= \int u^2\big(1+u^2\big)^2\;du.$$

We leave the integration and subsequent substitution to the reader. The final answer is

$$= \frac{1}{3}\tan^3(x)+\frac{2}{5}\tan^5(x)+\frac{1}{7}\tan^7(x)+C.$$

2. We employ rule #4 of the workflow outlined above.

$$\begin{aligned}\int \tan^6(x)\;dx &= \int \tan^4(x)\tan^2(x)\;dx\\ &= \int \tan^4(x)\big(\sec^2(x)-1\big)\;dx\\ &= \int \tan^4(x)\sec^2(x)\;dx-\int \tan^4(x)\;dx\end{aligned}$$

Integrate the first integral with substitution, $u=\tan(x)$; integrate the second by employing rule #4 again.

$$\begin{aligned}&= \frac{1}{5}\tan^5(x)-\int \tan^2(x)\tan^2(x)\;dx\\ &= \frac{1}{5}\tan^5(x)-\int \tan^2(x)\big(\sec^2(x)-1\big)\;dx\\ &= \frac{1}{5}\tan^5(x)-\int \tan^2(x)\sec^2(x)\;dx+\int \tan^2(x)\;dx\end{aligned}$$

Again, use substitution for the first integral and rule #4 for the second.

$$= \frac{1}{5}\tan^5(x)-\frac{1}{3}\tan^3(x)+\int \big(\sec^2(x)-1\big)\;dx$$

$$= \frac{1}{5}\tan^5(x) - \frac{1}{3}\tan^3(x) + \tan(x) - x + C$$

These latter examples were admittedly long, with repeated applications of the same rule. Try to not be overwhelmed by the length of the problem, but rather admire how robust this solution method is. A trigonometric function of a high power can be systematically reduced to trigonometric functions of lower powers until all antiderivatives can be computed.

12.4.4 Trigonometric substitution

We have since learned a number of integration techniques, yet we are still unable to evaluate an integral like

$$\int_{-3}^{3} \sqrt{9-x^2}\; dx. \tag{12.21}$$

without resorting to a geometric interpretation. This section introduces **trigonometric substitution** (*goniometrische substitutie*), a method of integration that fills this gap in our integration skill. This technique works on the same principle as substitution, by setting set $x = f(\theta)$, where f is a trigonometric function, and then replacing x with $f(\theta)$.

For what concerns the integral given by Equation (12.21), we begin by noting that $9\sin^2(\theta) + 9\cos^2(\theta) = 9$, and hence $9\cos^2(\theta) = 9 - 9\sin^2(\theta)$. If we let $x = 3\sin(\theta)$, then $9 - x^2 = 9 - 9\sin^2(\theta) = 9\cos^2(\theta)$.

Setting $x = 3\sin(\theta)$ gives $dx = 3\cos(\theta)\; d\theta$. We are almost ready to substitute. We also wish to change our bounds of integration. The bound $x = -3$ corresponds to $\theta = -\pi/2$. Likewise, the bound of $x = 3$ is replaced by the bound $\theta = \pi/2$. Thus

$$\begin{aligned}
\int_{-3}^{3} \sqrt{9-x^2}\; dx &= \int_{-\pi/2}^{\pi/2} \sqrt{9-9\sin^2(\theta)}(3\cos(\theta))\; d\theta \\
&= \int_{-\pi/2}^{\pi/2} 3\sqrt{9\cos^2(\theta)}\cos(\theta)\; d\theta \\
&= \int_{-\pi/2}^{\pi/2} 3\left|3\cos(\theta)\right|\cos(\theta)\; d\theta.
\end{aligned}$$

On $[-\pi/2, \pi/2]$, $\cos\theta$ is always positive, so we can drop the absolute value bars, then employ a power–reducing formula:

$$\begin{aligned}
&= \int_{-\pi/2}^{\pi/2} 9\cos^2(\theta)\; d\theta \\
&= \int_{-\pi/2}^{\pi/2} \frac{9}{2}\big(1+\cos(2\theta)\big)\; d\theta
\end{aligned}$$

$$= \frac{9}{2}\left(\theta + \frac{1}{2}\sin(2\theta)\right)\Bigg|_{-\pi/2}^{\pi/2} = \frac{9}{2}\pi.$$

This matches our answer in Example 12.3.

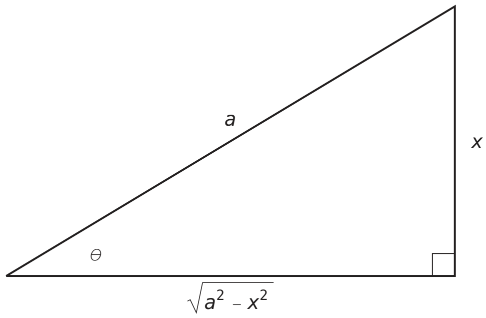
Trigonometric substitution excels when dealing with integrands that contain $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ and $\sqrt{x^2+a^2}$. The following outlines the procedure for each case. Each right triangle acts as a reference to help us understand the relationships between x and θ.

(a) For integrands containing $\sqrt{a^2-x^2}$:

Let $x = a\sin(\theta)$, then $dx = a\cos(\theta)\ d\theta$.

Thus $\theta = \arcsin(x/a)$, for $-\pi/2 \leq \theta \leq \pi/2$.

On this interval, $\cos(\theta) \geq 0$, so $\sqrt{a^2-x^2} = a\cos(\theta)$.

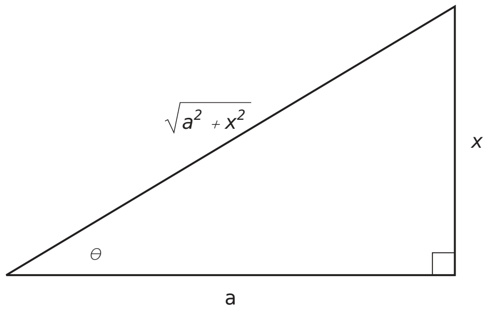


(b) For integrands containing $\sqrt{x^2+a^2}$:

Let $x = a\tan(\theta)$, then $dx = a\sec^2(\theta)\ d\theta$.

Thus $\theta = \arctan(x/a)$, for $-\pi/2 < \theta < \pi/2$.

On this interval, $\sec(\theta) > 0$, so $\sqrt{x^2+a^2} = a\sec(\theta)$.

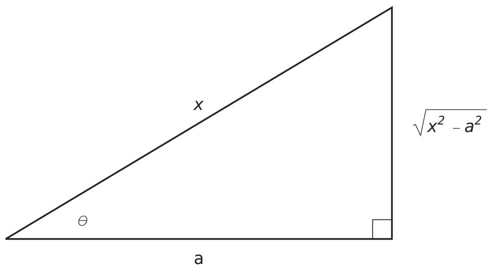


(c) For integrands containing $\sqrt{x^2-a^2}$:

Let $x = a\sec(\theta)$, then $dx = a\sec(\theta)\tan(\theta)\ d\theta$.

Thus $\theta = \operatorname{arcsec}(x/a)$. If $x/a \geq 1$, then $0 \leq \theta < \pi/2$; if $x/a \leq -1$, then $\pi/2 < \theta \leq \pi$.

We restrict our work to where $x \geq a$, so $x/a \geq 1$, and $0 \leq \theta < \pi/2$. On this interval, $\tan\theta \geq 0$, so $\sqrt{x^2-a^2} = a\tan(\theta)$.



Example 12.24

Evaluate

$$\int \sqrt{4x^2-1}\ dx.$$

Solution

We start by rewriting the integrand so that it looks like $\sqrt{x^2-a^2}$ for some value of a:

$$\sqrt{4x^2-1} = \sqrt{4\left(x^2 - \frac{1}{4}\right)}$$

$$= 2\sqrt{x^2 - \left(\frac{1}{2}\right)^2}.$$

So we have $a = 1/2$, and following rule (c) from the above workflow, we set $x = \sec(\theta)/2$, and hence $dx = \sec(\theta)\tan(\theta)/2\ d\theta$. We now rewrite the integral with these substitutions:

$$\begin{aligned}
\int \sqrt{4x^2-1}\ dx &= \int 2\sqrt{x^2 - \left(\frac{1}{2}\right)^2}\ dx \\
&= \int 2\sqrt{\frac{1}{4}\sec^2(\theta) - \frac{1}{4}}\left(\frac{1}{2}\sec(\theta)\tan(\theta)\right)\ d\theta \\
&= \int \sqrt{\frac{1}{4}(\sec^2(\theta)-1)}\left(\sec(\theta)\tan(\theta)\right)\ d\theta \\
&= \int \sqrt{\frac{1}{4}\tan^2(\theta)}\left(\sec\theta\tan(\theta)\right)\ d\theta \\
&= \int \frac{1}{2}\tan^2(\theta)\sec(\theta)\ d\theta \\
&= \frac{1}{2}\int\left(\sec^2(\theta)-1\right)\sec(\theta)\ d\theta \\
&= \frac{1}{2}\int\left(\sec^3(\theta)-\sec(\theta)\right)\ d\theta.
\end{aligned}$$

We can now integrate $\sec^3(\theta)$ using integration by parts with $dv = \sec^2(\theta)$ and $u = \sec(\theta)$, finding its antiderivatives to be

$$\int \sec^3(\theta)\ d\theta = \frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\left|\sec(\theta)+\tan(\theta)\right|\Big) + C.$$

Thus

$$\begin{aligned}
\int \sqrt{4x^2-1}\ dx &= \frac{1}{2}\int\left(\sec^3(\theta)-\sec(\theta)\right)\ d\theta \\
&= \frac{1}{2}\left(\frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln\left|\sec(\theta)+\tan(\theta)\right|\Big) - \ln\left|\sec(\theta)+\tan(\theta)\right|\right) + C \\
&= \frac{1}{4}\left(\sec(\theta)\tan(\theta) - \ln\left|\sec(\theta)+\tan(\theta)\right|\right) + C.
\end{aligned}$$

We are not yet done. Our original integral is given in terms of x, whereas our final answer, as given, is in terms of θ. We need to rewrite our answer in terms of x. With $a = 1/2$, and $x = \sec(\theta)/2$, the reference triangle in rule (c) of the above workflow shows that

$$\tan\theta = \sqrt{x^2 - \frac{1}{4}}\Big/\frac{1}{2} = 2\sqrt{x^2-\frac{1}{4}} \quad \text{and} \quad \sec(\theta) = 2x.$$

Thus

$$\frac{1}{4}\Big(\sec(\theta)\tan(\theta) - \ln\left|\sec(\theta)+\tan(\theta)\right|\Big) + C$$

becomes

$$\frac{1}{4}\left(2x\,2\sqrt{x^2-\frac{1}{4}}-\ln\left|2x+2\sqrt{x^2-\frac{1}{4}}\right|\right)+C.$$

The final answer hence is:

$$\int\sqrt{4x^2-1}\,dx=\frac{1}{4}\left(4x\sqrt{x^2-\frac{1}{4}}-\ln\left|2x+2\sqrt{x^2-\frac{1}{4}}\right|\right)+C.$$

It is important to realize that trigonometric substitution can be applied in many situations, even those not of the form $\sqrt{a^2-x^2}$, $\sqrt{x^2-a^2}$ or $\sqrt{x^2+a^2}$. This is illustrated in the following example.

Example 12.25

Evaluate

$$\int\frac{1}{(x^2+6x+10)^2}\,dx.$$

Solution

We start by completing the square, then make the substitution $u=x+3$, followed by the trigonometric substitution of $u=\tan(\theta)$:

$$\int\frac{1}{(x^2+6x+10)^2}\,dx=\int\frac{1}{\left((x+3)^2+1\right)^2}\,dx=\int\frac{1}{(u^2+1)^2}\,du.$$

Now make the substitution $u=\tan(\theta)$, $du=\sec^2(\theta)\,d\theta$:

$$\begin{aligned}\int\frac{1}{(u^2+1)^2}\,du&=\int\frac{1}{(\tan^2(\theta)+1)^2}\sec^2(\theta)\,d\theta\\&=\int\frac{1}{(\sec^2(\theta))^2}\sec^2(\theta)\,d\theta\\&=\int\cos^2(\theta)\,d\theta.\end{aligned}$$

Applying a power reducing formula, we have

$$\begin{aligned}\int\cos^2(\theta)\,d\theta&=\int\left(\frac{1}{2}+\frac{1}{2}\cos(2\theta)\right)\,d\theta\\&=\frac{1}{2}\theta+\frac{1}{4}\sin(2\theta)+C.\end{aligned}\tag{12.22}$$

We need to return to the variable x. As $u=\tan(\theta)$, $\theta=\arctan(u)$. Using the identity $\sin(2\theta)=2\sin(\theta)\cos(\theta)$ and using the reference triangle found in rule (b) of the workflow above, we have

$$\frac{1}{4}\sin(2\theta)=\frac{1}{2}\frac{u}{\sqrt{u^2+1}}\frac{1}{\sqrt{u^2+1}}=\frac{1}{2}\frac{u}{u^2+1}.$$

Finally, we return to x with the substitution $u=x+3$. We start with the expression in Equation (12.22):

$$\begin{aligned}\frac{1}{2}\theta+\frac{1}{4}\sin(2\theta)+C &= \frac{1}{2}\arctan(u)+\frac{1}{2}\frac{u}{u^2+1}+C\\ &= \frac{1}{2}\arctan(x+3)+\frac{x+3}{2(x^2+6x+10)}+C.\end{aligned}$$

Stating our final result in one line:

$$\int \frac{1}{(x^2+6x+10)^2}\,dx = \frac{1}{2}\arctan(x+3)+\frac{x+3}{2(x^2+6x+10)}+C.$$

Finally, it should be mentioned that given a definite integral that can be evaluated using trigonometric substitution, we could first evaluate the corresponding indefinite integral and then evaluate using the original bounds. It is much more straightforward, though, to change the bounds as we substitute.

12.4.5 Partial fraction decomposition

Here we investigate the antiderivatives of rational functions. Recall that rational functions are functions of the form $f(x)=\frac{p(x)}{q(x)}$, where $p(x)$ and $q(x)$ are polynomials and $q(x)\neq 0$.

Consider the integral

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

We do not have a simple formula for this. It can be evaluated using trigonometric substitution, but note how the integral is easy to evaluate once we realize:

$$\frac{1}{x^2-1}=\frac{1/2}{x-1}-\frac{1/2}{x+1}.$$

Thus

$$\begin{aligned}\int \frac{1}{x^2-1}\,dx &= \int \frac{1/2}{x-1}\,dx-\int \frac{1/2}{x+1}\,dx\\ &= \frac{1}{2}\ln|x-1|-\frac{1}{2}\ln|x+1|+C.\end{aligned}$$

Here, we will learn how to decompose fractions like

$$\frac{1}{x^2-1}.$$

We start with a rational function

$$f(x)=\frac{p(x)}{q(x)},$$

where p and q do not have any common factors and the degree of p is less than the degree of q. It can be shown that any polynomial, and hence q, can be factored into a product of real linear and irreducible quadratic terms. The following workflow states how to **decompose a rational function into partial fractions** (*splitsing in partieelbreuken*) as a sum of rational functions whose denominators are all of lower degree than q.

1. **Linear Terms:** Let $(x-a)$ divide $q(x)$, where $(x-a)^n$ is the highest power of $(x-a)$ that divides

$q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{A_1}{(x-a)}+\frac{A_2}{(x-a)^2}+\cdots+\frac{A_n}{(x-a)^n}.$$

2. **Quadratic Terms:** Let (x^2+bx+c) divide $q(x)$, where $(x^2+bx+c)^n$ is the highest power of (x^2+bx+c) that divides $q(x)$. Then the decomposition of $f(x)$ will contain the sum

$$\frac{B_1x+C_1}{x^2+bx+c}+\frac{B_2x+C_2}{(x^2+bx+c)^2}+\cdots+\frac{B_nx+C_n}{(x^2+bx+c)^n}.$$

To find the coefficients A_i, B_i and C_i:

1. Multiply all fractions by $q(x)$, clearing the denominators. Collect like terms.
2. Equate the resulting coefficients of the powers of x and solve the resulting system of linear equations.

Example 12.26

Perform the partial fraction decomposition of

$$\frac{1}{x^2-1}.$$

Solution

The denominator factors into two linear terms: $x^2-1=(x-1)(x+1)$. Thus

$$\frac{1}{x^2-1}=\frac{A}{x-1}+\frac{B}{x+1}.$$

To solve for A and B, first multiply through by $x^2-1=(x-1)(x+1)$:

$$\begin{aligned}1&=\frac{A(x-1)(x+1)}{x-1}+\frac{B(x-1)(x+1)}{x+1}\\&=A(x+1)+B(x-1)\\&=Ax+A+Bx-B\\&=(A+B)x+(A-B).\end{aligned}$$

The next step is key. Note the equality we have:

$$1=(A+B)x+(A-B).$$

For clarity's sake, rewrite the left hand side as

$$0x+1=(A+B)x+(A-B).$$

On the left, the coefficient of the x term is 0; on the right, it is $(A+B)$. Since both sides are equal, we must have that $0=A+B$.

Likewise, on the left, we have a constant term of 1; on the right, the constant term is $(A-B)$. Therefore we have $1=A-B$.

We have two linear equations with two unknowns. This one is easy to solve by hand, leading to

$$\begin{cases} A+B=0 \\ A-B=1 \end{cases} \quad \Rightarrow \quad \begin{cases} A=1/2 \\ B=-1/2. \end{cases}$$

Thus

$$\frac{1}{x^2-1} = \frac{1/2}{x-1} - \frac{1/2}{x+1}.$$

Clearly, it can become rather tedious to do a a partial fraction decomposition by hand if one is confronted with a more complex rational fraction. Luckily, we can resort in such cases to Mathematica, which can accomplish this with the command **`Apart`**. For instance, for what concerns the rational function in Example (12.26), we should proceed as follows.

```
In[21]:= Apart[1/(x^2 - 1), x]
```

The second argument of the command **`Apart`** is nothing but the variable at stake.

```
Out[21]=     1           1
          ---------- - ---------
          2 (-1+x)     2 (1+x)
```

Example 12.27

Evaluate the following indefinite integrals:

1. $\displaystyle\int \frac{1}{(x-1)(x+2)^2}\,dx,$
2. $\displaystyle\int \frac{x^3}{(x-5)(x+3)}\,dx,$
3. $\displaystyle\int \frac{2+\sin(x)}{3+\cos(x)}\,dx.$

Solution

1. We decompose the integrand as follows:

$$\frac{1}{(x-1)(x+2)^2} = \frac{A}{x-1} + \frac{B}{x+2} + \frac{C}{(x+2)^2}.$$

To solve for A, B and C, we multiply both sides by $(x-1)(x+2)^2$ and collect like terms:

$$\begin{aligned} 1 &= A(x+2)^2 + B(x-1)(x+2) + C(x-1) \\ &= Ax^2 + 4Ax + 4A + Bx^2 + Bx - 2B + Cx - C \\ &= (A+B)x^2 + (4A+B+C)x + (4A-2B-C). \end{aligned} \tag{12.23}$$

We have

$$0x^2 + 0x + 1 = (A+B)x^2 + (4A+B+C)x + (4A-2B-C),$$

leading to the equations

$$\begin{cases} A+B &= 0 \\ 4A+B+C &= 0 \\ 4A-2B-C &= 1 \end{cases} \quad \Leftrightarrow \quad \begin{cases} A &= \frac{1}{9} \\ B &= -\frac{1}{9} \\ C &= -\frac{1}{3} \end{cases}$$

Thus

$$\int \frac{1}{(x-1)(x+2)^2}\,dx = \int \frac{1/9}{x-1}\,dx + \int \frac{-1/9}{x+2}\,dx + \int \frac{-1/3}{(x+2)^2}\,dx.$$

Each can be integrated with a simple substitution with $u = x-1$ or $u = x+2$. The end result is

$$\int \frac{1}{(x-1)(x+2)^2}\,dx = \frac{1}{9}\ln|x-1| - \frac{1}{9}\ln|x+2| + \frac{1}{3(x+2)} + C.$$

2. Since the degree of the numerator is now higher than the one of the denominator, we begin by using polynomial division to reduce the degree of the numerator (see Section 4.1). Doing so, we arrive at

$$\frac{x^3}{(x-5)(x+3)} = x+2+\frac{19x+30}{(x-5)(x+3)}.$$

Consequently, we can rewrite the new rational function as:

$$\frac{19x+30}{(x-5)(x+3)} = \frac{A}{x-5} + \frac{B}{x+3},$$

for appropriate values of A and B. Clearing denominators, we have

$$\begin{aligned} 19x+30 &= A(x+3)+B(x-5) \\ &= (A+B)x+(3A-5B). \end{aligned}$$

This implies that:

$$\begin{cases} 19 = A+B \\ 30 = 3A-5B. \end{cases} \quad \Leftrightarrow \quad \begin{cases} A = \frac{125}{8} \\ B = \frac{27}{8}. \end{cases}$$

We can now integrate:

$$\begin{aligned} \int \frac{x^3}{(x-5)(x+3)}\,dx &= \int \left(x+2+\frac{125/8}{x-5}+\frac{27/8}{x+3}\right)dx \\ &= \frac{x^2}{2}+2x+\frac{125}{8}\ln|x-5|+\frac{27}{8}\ln|x+3|+C. \end{aligned}$$

3. We observe that we are confronted with a rational function of trigonometric functions, so we first of all resort to the Weierstrass substitution. This leads to the following integral

$$2\int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\,dt,$$

which can be finished off using partial fraction decomposition. In this way, we get

$$2\int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\,dt = 2\int \frac{1-t}{t^2+2}\,dt + \int \frac{t}{t^2+1}\,dt\,.$$

Hence, we arrive at

$$2\int \frac{t^2+t+1}{(t^2+2)(t^2+1)}\,dt = \frac{1}{\sqrt{2}}\arctan\left(\frac{t}{\sqrt{2}}\right) - \frac{1}{2}\ln\left(t^2+2\right) + \frac{1}{2}\ln\left(t^2+1\right) + C\,,$$

where $t = \tan(x/2)$.

We conclude our discussion of partial fraction decomposition with a final example that combines several of the techniques we encountered earlier in this section.

Example 12.28

Evaluate

$$\int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\,dx.$$

Solution

The degree of the numerator is less than the degree of the denominator, so we have:

$$\frac{7x^2+31x+54}{(x+1)(x^2+6x+11)} = \frac{A}{x+1} + \frac{Bx+C}{x^2+6x+11}.$$

Now clear the denominators.

$$\begin{aligned} 7x^2+31x+54 &= A(x^2+6x+11)+(Bx+C)(x+1) \\ &= (A+B)x^2+(6A+B+C)x+(11A+C). \end{aligned}$$

This implies that:

$$\begin{cases} 7 = A+B \\ 31 = 6A+B+C \\ 54 = 11A+C. \end{cases} \quad \Leftrightarrow \quad \begin{cases} A = 5 \\ B = 2 \\ C = -1. \end{cases}$$

Thus

$$\int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\,dx = \int \left(\frac{5}{x+1} + \frac{2x-1}{x^2+6x+11}\right)dx.$$

The first term of this new integrand is easy to evaluate; it leads to a $5\ln|x+1|$ term. The second term is not hard, but takes several steps and uses substitution techniques.

The integrand $\dfrac{2x-1}{x^2+6x+11}$ has a quadratic in the denominator and a linear term in the numerator. This leads us to try substitution. Let $u = x^2+6x+11$, so $du = (2x+6)\,dx$. The numerator is $2x-1$, not $2x+6$, but we can get a $2x+6$ term in the numerator by adding 0 in the form of "$7-7$."

$$\begin{aligned} \frac{2x-1}{x^2+6x+11} &= \frac{2x-1+7-7}{x^2+6x+11} \\ &= \frac{2x+6}{x^2+6x+11} - \frac{7}{x^2+6x+11}. \end{aligned}$$

We can now integrate the first term with substitution, leading to a $\ln|x^2+6x+11|$ term. The final term can be integrated using arctangent. First, complete the square in the denominator:

$$\frac{7}{x^2+6x+11} = \frac{7}{(x+3)^2+2}.$$

An antiderivative of the latter term can be found using Equation (12.13) and substitution:

$$\int \frac{7}{x^2+6x+11}\,dx = \frac{7}{\sqrt{2}}\arctan\left(\frac{x+3}{\sqrt{2}}\right)+C.$$

Let's start at the beginning and put all of the steps together.

$$\begin{aligned}\int \frac{7x^2+31x+54}{(x+1)(x^2+6x+11)}\,dx &= \int\left(\frac{5}{x+1}+\frac{2x-1}{x^2+6x+11}\right)dx \\ &= \int \frac{5}{x+1}\,dx + \int \frac{2x+6}{x^2+6x+11}\,dx - \int \frac{7}{x^2+6x+11}\,dx \\ &= 5\ln|x+1| + \ln\left|x^2+6x+11\right| - \frac{7}{\sqrt{2}}\arctan\left(\frac{x+3}{\sqrt{2}}\right)+C\end{aligned}$$

It is important to remember that one is not expected to see the final answer immediately after seeing the problem. Rather, given the initial problem, we break it down into smaller problems that are easier to solve. The final answer is a combination of the answers of the smaller problems.

Partial fraction decomposition is an important tool when dealing with rational functions. Note that at its heart, it is a technique of algebra, not calculus, as we are rewriting a fraction in a new form. Still, it is very useful in the realm of calculus as it lets us evaluate a certain set of complicated integrals.

Integral equations

In Chapter 9, we encountered differential equations, which are equations that relate some function with its derivatives. Likewise, we can formulate integral equations, which are equations in which an unknown function appears under an integral sign. Consider, for instance, the following integral equation:

$$f(x) = e^{-x} - \frac{1}{2} + \frac{1}{2}e^{-x-1} + \frac{1}{2}\int_0^1 (x+1)e^{-xy}f(y)\,dy.$$

Its solution is $f(x) = e^{-x}$, which can verified easily.

Just are the differential equations, integral equations are omnipresent in physics and engineering. For instance, Maxwell's equations of electromagnetism can be formulated in integral form.

12.5 Improper Integration

Consider the following definite integrals:

- $\displaystyle\int_0^{100} \frac{1}{1+x^2}\,dx \approx 1.5608,$
- $\displaystyle\int_0^{1000} \frac{1}{1+x^2}\,dx \approx 1.5698,$
- $\displaystyle\int_0^{10,000} \frac{1}{1+x^2}\,dx \approx 1.5707.$

Notice how the integrand is $1/(1+x^2)$ in each integral. It is sketched in Figure 12.15. As the upper bound gets larger, one would expect the area under the curve would also grow. While the definite integrals do increase in value as the upper bound grows, they are not increasing by much. In fact, consider:

$$\int_0^b \frac{1}{1+x^2}\,dx = \arctan(x)\Big|_0^b = \arctan(b) - \arctan(0) = \arctan(b).$$

As $b \to +\infty$, $\arctan(b) \to \pi/2$. Therefore it seems that as the upper bound b grows, the value of the concerned definite integral approaches $\pi/2 \approx 1.5708$. This should strike the reader as being a bit amazing: even though the curve extends to infinity, it has a finite amount of area underneath it.

When we defined the definite integral $\int_a^b f(x)\,dx$ in Definition 12.2, we made two stipulations:

1. The interval over which we integrated, $[a,b]$, was a finite interval, and
2. The function $f(x)$ was continuous on $[a,b]$ (ensuring that the range of f was finite).

In this section we consider integrals where one or both of the above conditions do not hold. Such integrals are called **improper integrals** (*oneigenlijke integraal*)

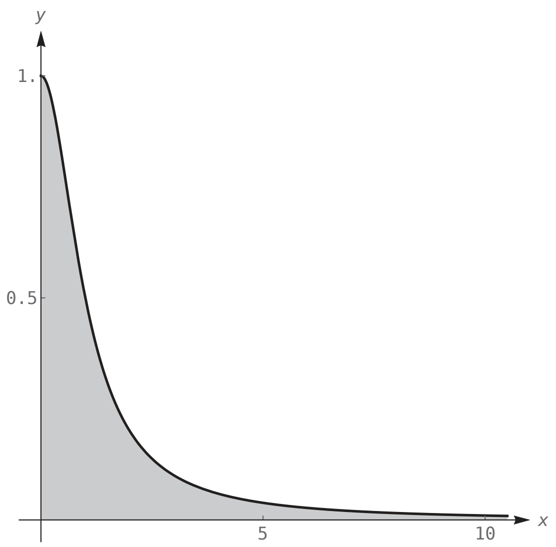


Figure 12.15: Graphing $f(x) = \dfrac{1}{1+x^2}$.

12.5.1 Improper integrals with infinite bounds

We start with a definition of Improper integrals with infinite bounds.

Definition 12.6 (Improper integrals with infinite bounds)

1. Let f be a continuous function on $[a,+\infty[$. Define

$$\int_a^{+\infty} f(x)\,dx \quad \text{to be} \quad \lim_{b\to+\infty}\int_a^b f(x)\,dx.$$

2. Let f be a continuous function on $]-\infty, b]$. Define

$$\int_{-\infty}^{b} f(x)\, dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{b} f(x)\, dx.$$

3. Let f be a continuous function on $\mathbb{R}$. Let c be any real number; define

$$\int_{-\infty}^{+\infty} f(x)\, dx \quad \text{to be} \quad \lim_{a\to-\infty} \int_{a}^{c} f(x)\, dx + \lim_{b\to+\infty} \int_{c}^{b} f(x)\, dx.$$

An improper integral is said to converge if its corresponding limit exists (is finite); otherwise, it diverges. The improper integral in part 3 converges if and only if both of its limits exist.

Example 12.29

Evaluate the following improper integrals:

1. $\displaystyle\int_{1}^{+\infty} \frac{1}{x^2}\, dx,$
2. $\displaystyle\int_{1}^{+\infty} \frac{1}{x}\, dx,$
3. $\displaystyle\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\, dx.$

Solution

1.

$$[t] \int_{1}^{+\infty} \frac{1}{x^2}\, dx = \lim_{b\to+\infty} \int_{1}^{b} \frac{1}{x^2}\, dx = \lim_{b\to+\infty} \left.\frac{-1}{x}\right|_{1}^{b} \tag{12.24}$$

$$= \lim_{b\to+\infty} \frac{-1}{b} + 1 = 1. \tag{12.25}$$

A graph of the area defined by this integral is given in Figure 12.16(a). In Mathematica, this result can be checked as follows:

```
In[22]:= Integrate[1/x^2, x, 1, +Infinity]

Out[22]= 1
```

2.

$$\begin{aligned}\int_{1}^{\infty} \frac{1}{x}\, dx &= \lim_{b\to+\infty} \int_{1}^{b} \frac{1}{x}\, dx \\ &= \lim_{b\to+\infty} \ln|x|\Big|_{1}^{b} \\ &= \lim_{b\to+\infty} \ln(b) \\ &= +\infty.\end{aligned}$$

The limit does not exist, hence the concerned improper integral diverges. Compare the graphs in Figures 12.16(a) and 12.16(b); notice how the graph of $f(x) = 1/x$ is noticeably larger. This difference is enough to cause the improper integral to diverge.

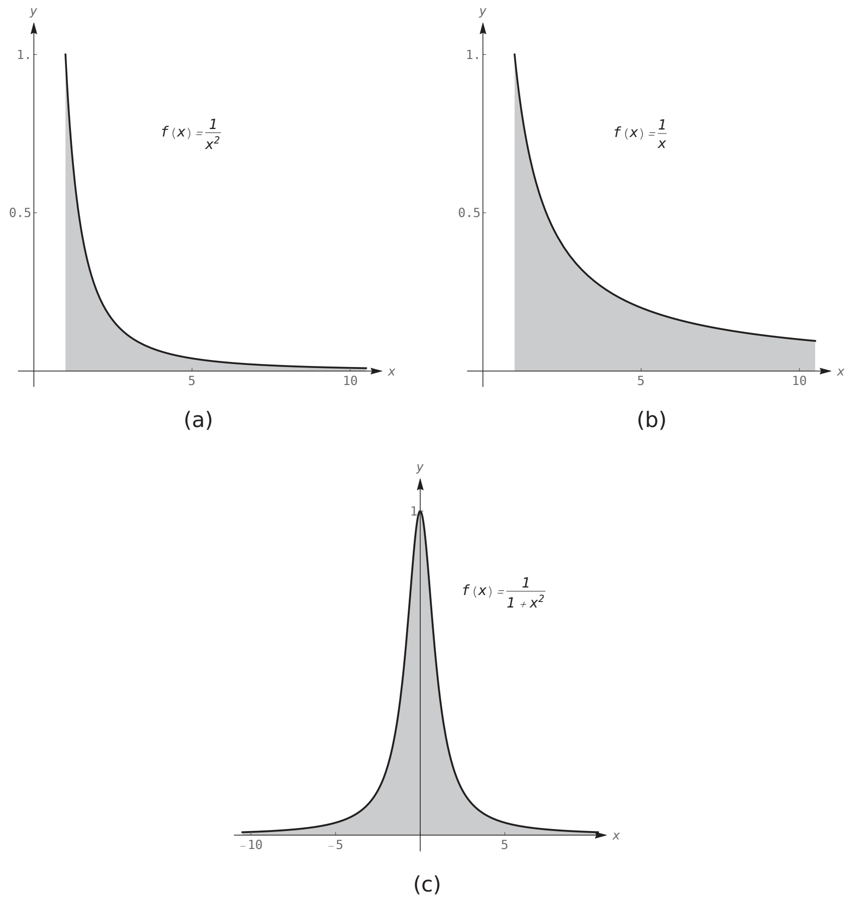


Figure 12.16: A graph of $f(x) = \frac{1}{x^2}$ (a), $f(x) = \frac{1}{x}$ (b) and $f(x) = \frac{1}{1+x^2}$ (c) in Example 12.29.

3. We will need to break this into two improper integrals and choose a value of c as in part 3 of Definition 12.6. Any value of c is fine; we choose $c = 0$.

$$\begin{aligned}
\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx &= \lim_{a\to-\infty} \int_a^0 \frac{1}{1+x^2}\,dx + \lim_{b\to+\infty} \int_0^b \frac{1}{1+x^2}\,dx \\
&= \lim_{a\to-\infty} \arctan(x)\Big|_a^0 + \lim_{b\to+\infty} \arctan(x)\Big|_0^b \\
&= \lim_{a\to-\infty} \big(\arctan(0) - \arctan(a)\big) + \lim_{b\to+\infty} \big(\arctan(b) - \arctan(0)\big) \\
&= \left(0 - \frac{-\pi}{2}\right) + \left(\frac{\pi}{2} - 0\right)
\end{aligned}$$

Each limit exists, hence the original integral converges and has value:

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx = \pi.$$

A graph of the area defined by this integral is given in Figure 12.16(c).

Note that it is not uncommon for the limits resulting from improper integrals to need l'Hôpital's rule.

12.5.2 Improper integrals with infinite range

We have just considered definite integrals where the interval of integration was infinite. We now consider another type of improper integration, where the range of the integrand is infinite.

Definition 12.7 (Improper integrals with infinite range)
Let $f(x)$ be a continuous function on $[a,b]$ except at c, $a \le c \le b$, where $x = c$ is a vertical asymptote of f. Define

$$\int_a^b f(x)\,dx = \lim_{\substack{t\to c\\ <}} \int_a^t f(x)\,dx + \lim_{\substack{t\to c\\ >}} \int_t^b f(x)\,dx.$$

Example 12.30

Evaluate the following improper integrals:

1. $\displaystyle\int_0^1 \frac{1}{\sqrt{x}}\,dx,$

2. $\displaystyle\int_{-1}^1 \frac{1}{x^2}\,dx.$

Solution

1. A graph of $f(x) = 1/\sqrt{x}$ is given in Figure 12.17(a). Notice that f has a vertical asymptote at $x = 0$; in some sense, we are trying to compute the area of a region that has no top. Could this have a finite value?

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{x}}\,dx &= \lim_{\substack{a\to 0\\ >}} \int_a^1 \frac{1}{\sqrt{x}}\,dx \\ &= \lim_{\substack{a\to 0\\ >}} 2\sqrt{x}\Big|_a^1 \\ &= \lim_{\substack{a\to 0\\ >}} 2\left(\sqrt{1} - \sqrt{a}\right) \\ &= 2\end{aligned}$$

It turns out that the region does have a finite area even though it has no upper bound.

2. The function $f(x) = 1/x^2$ has a vertical asymptote at $x = 0$, as shown in Figure 12.17(b), so this integral is an improper integral. Let's eschew using limits for a moment and proceed without recognizing the improper nature of the integral. This leads to:

$$\int_{-1}^1 \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_{-1}^1$$

$$= -1 - (1)$$
$$= -2. \ (!)$$

Clearly the area in question is above the x-axis, yet the area is supposedly negative! Why does our answer not match our intuition? To answer this, evaluate the integral using Definition 12.7.

$$\int_{-1}^{1} \frac{1}{x^2}\, dx = \lim_{\substack{t\to 0\\ <}} \int_{-1}^{t} \frac{1}{x^2}\, dx + \lim_{\substack{t\to 0\\ >}} \int_{t}^{1} \frac{1}{x^2}\, dx$$
$$= \lim_{\substack{t\to 0\\ <}} \left(-\frac{1}{x}\right)\Big|_{-1}^{t} + \lim_{\substack{t\to 0\\ >}} \left(-\frac{1}{x}\right)\Big|_{t}^{1}$$
$$= \lim_{\substack{t\to 0\\ <}} \left(-\frac{1}{t} - 1\right) + \lim_{\substack{t\to 0\\ >}} \left(-1 + \frac{1}{t}\right)$$
$$= \Big(+\infty - 1\Big) + \Big(-1 + \infty\Big)$$

Neither limit converges hence the original improper integral diverges. The nonsensical answer we obtained by ignoring the improper nature of the integral is just that: nonsensical.

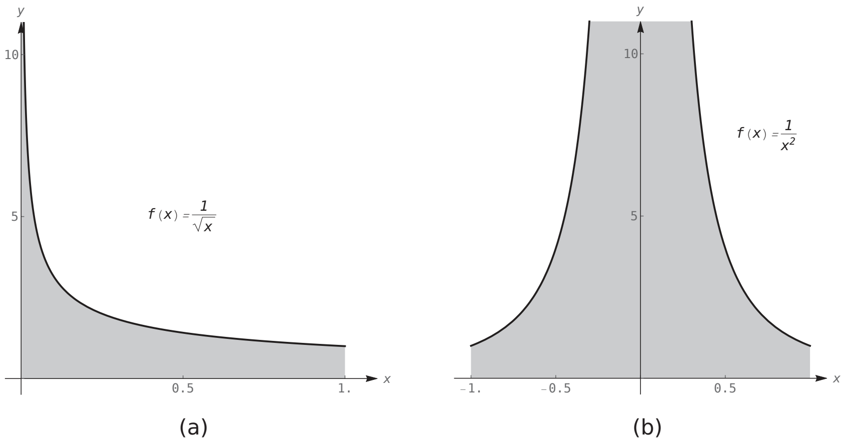


Figure 12.17: A graph of $f(x) = \frac{1}{\sqrt{x}}$ (a) and $f(x) = \frac{1}{x^2}$ (b) in Example 12.30.

12.5.3 Convergence and divergence

Oftentimes we are interested in knowing simply whether or not an improper integral converges, and not necessarily the value of a convergent integral. We provide here several tools that help determine the **convergence** (*convergentie*) or **divergence** (*divergentie*) of improper integrals without integrating.

For instance, let us try to determine the values of p for which

$$\int_{1}^{+\infty} \frac{1}{x^p}\, dx$$

converges.

We begin by integrating and then evaluating the limit:

$$\begin{aligned}
\int_1^{+\infty} \frac{1}{x^p}\,dx &= \lim_{b\to+\infty} \int_1^b \frac{1}{x^p}\,dx \\
&= \lim_{b\to+\infty} \int_1^b x^{-p}\,dx \qquad \text{(assume } p \neq 1\text{)} \\
&= \lim_{b\to+\infty} \frac{1}{-p+1} x^{-p+1}\Big|_1^b \\
&= \lim_{b\to+\infty} \frac{1}{1-p}\big(b^{1-p} - 1^{1-p}\big).
\end{aligned}$$

When does this limit converge – i.e., when is this limit not ∞? This limit converges precisely when the power of b is less than 0: when $1-p<0 \Rightarrow 1<p$.

So, if $p>1$, then $\int_1^\infty \frac{1}{x^p}\,dx$ converges. When $p<1$ the improper integral diverges; we showed in Example 12.29 that when $p=1$ the integral also diverges. Figure 12.18 graphs $y=1/x$ with a dashed line, along with graphs of $y=1/x^p$, $p<1$, and $y=1/x^q$, $q>1$. Somehow the dashed line forms a dividing line between convergence and divergence.

A similar result is proved in the exercises about improper integrals of the form

$$\int_0^1 \frac{1}{x^p}\,dx,$$

i.e. this improper integral converges when $p<1$ and diverges when $p\geq 1$.

Note that we used the upper and lower bound of 1 just for convenience. It can be replaced by any a where $a>0$.

A basic technique in determining convergence of improper integrals is to compare an integrand whose convergence is unknown to an integrand whose convergence is known. We often use integrands of the form $1/x^p$ to compare to as their convergence on certain intervals is known. This is described in the following theorem.

Theorem 12.9 (Direct comparison test for improper integrals)
Let f and g be continuous on $[a,+\infty[$ where $0\leq f(x)\leq g(x)$ for all x in $[a,+\infty[$.

1. *If $\displaystyle\int_a^{+\infty} g(x)\,dx$ converges, then $\displaystyle\int_a^{+\infty} f(x)\,dx$ converges.*
2. *If $\displaystyle\int_a^{+\infty} f(x)\,dx$ diverges, then $\displaystyle\int_a^{+\infty} g(x)\,dx$ diverges.*

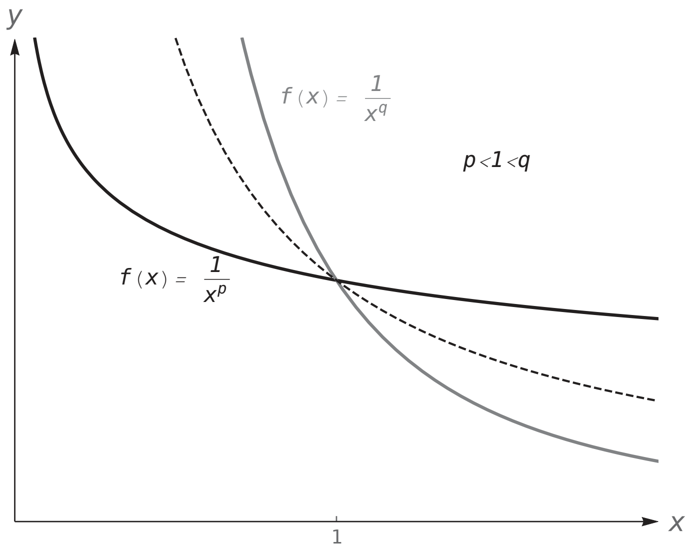


Figure 12.18: Plotting functions of the form $1/x^p$.

To prove Theorem 12.9, let us first of all prove the following theorem, which will need later on.

Theorem 12.10
Let $F(t)$ be an increasing function on an interval $]\,a, +\infty\,[$. Assume there exists $M > 0$ such that $F(t) \leq M$ for all $t \in\,]\,a, +\infty\,[$. Then the following limit exists:

$$L = \lim_{t \to +\infty} F(t),$$

and $L \leq M$.

Let S be the set of values of $F(t)$ on $]\,a, +\infty\,[$:

$$S = \{y \mid y = F(t), \, \forall t > a\}.$$

By assumption, S is bounded by M, that is, $y \leq M$ for all $y \in S$, so S has a least upper bound (supremum). Let $L = \sup(S)$. Then for all $\epsilon > 0$, $L - \epsilon$ is not an upper bound for S, so there exists some $y_0 > a$ such that $F(y_0) > L - \epsilon$. Since $F(t)$ is an increasing function, it follows that

$$L - \epsilon \leq F(y_0) \leq F(t) \leq L$$

for $t > y_0$. Therefore, $\left|L - F(t)\right| < \epsilon$ for $t > y_0$. Since ϵ is an arbitrary positive number, this is precisely what is needed to conclude that

$$L = \lim_{t \to +\infty} F(t).$$

Now to prove the first part of Theorem 12.9, consider the functions

$$G(t) = \int_a^t g(x)\, dx \qquad \text{and} \qquad F(t) = \int_a^t f(x)\, dx$$

They are defined for $t > a$. Since $f(x) \geq 0$ and $g(x) \geq 0$, both $F(t)$ and $G(t)$ are increasing. Furthermore, $f(x) \leq g(x)$ for all $x \geq a$ and therefore,

$$F(t) \leq G(t) \tag{12.26}$$

for all $t \geq a$.

Our assumption now is that the following improper integral converges:

$$M = \int_a^{+\infty} g(x)\ dx\,.$$

By definition, we have that $M = \lim_{t\to+\infty} G(t)$. Since $G(t)$ is increasing, it holds that

$$G(t) \leq M$$

for all $t \geq a$, and it subsequently follows from Inequality (12.26) that

$$F(t) \leq M$$

for all $t \geq a$.

Since we have shown that $F(t)$ is increasing and bounded by M, we can conclude that $\lim_{t\to+\infty} F(t)$ exists. Since this limit is equal to the desired improper integral

$$\lim_{t\to+\infty} F(t) = \int_a^{+\infty} f(x)\ dx\,,$$

this concludes our proof of the first part. Moreover, the second part follows immediately. Indeed, assume that the first part is known to be true and that

$$\int_a^{+\infty} f(x)\ dx$$

diverges. Then

$$\int_a^{+\infty} g(x)\ dx$$

must diverge as well, for if it converged, the first part would imply that

$$\int_a^{+\infty} f(x)\ dx$$

converges. Similarly, the second part implies the first.

There is also a counterpart of Theorem 12.9 for improper integrals with infinite range.

Theorem 12.11 (Direct comparison test for improper integrals with infinite range)
Let f and g be continuous on $[a, x_0[$ where $0 \leq f(x) \leq g(x)$ for all x in $[a, x_0[$.

1. *If $\displaystyle\int_a^{x_0} g(x)\ dx$ converges, then $\displaystyle\int_a^{x_0} f(x)\ dx$ converges.*
2. *If $\displaystyle\int_a^{x_0} f(x)\ dx$ diverges, then $\displaystyle\int_a^{x_0} g(x)\ dx$ diverges.*

Example 12.31

Determine the convergence of the following improper integrals.

1. $\displaystyle\int_1^{+\infty} e^{-x^2}\,dx$

2. $\displaystyle\int_3^{+\infty} \frac{1}{\sqrt{x^2-x}}\,dx$

Solution

1. The function $f(x) = e^{-x^2}$ does not have an antiderivative expressible in terms of elementary functions, so we cannot integrate directly. It is comparable to $g(x) = 1/x^2$, and as demonstrated in Figure 12.19(a), $e^{-x^2} < 1/x^2$ on $[1,+\infty[$. We know that $\int_1^{+\infty} x^{-2}\,dx$ converges, hence also the improper integral under consideration converges.

2. Note that for large values of x, we have

$$\frac{1}{\sqrt{x^2-x}} \approx \frac{1}{\sqrt{x^2}} = \frac{1}{x}.$$

We know that $\int_3^{+\infty} x^{-1}\,dx$ diverges, so we seek to compare the original integrand to $1/x$. It is easy to see that when $x > 0$, we have

$$x = \sqrt{x^2} > \sqrt{x^2-x} \quad\Leftrightarrow\quad \frac{1}{x} < \frac{1}{\sqrt{x^2-x}}.$$

Using Theorem 12.9, we conclude that since $\int_3^{+\infty} x^{-1}\,dx$ diverges, the concerned improper integral diverges as well. Figure 12.19(b) illustrates this.

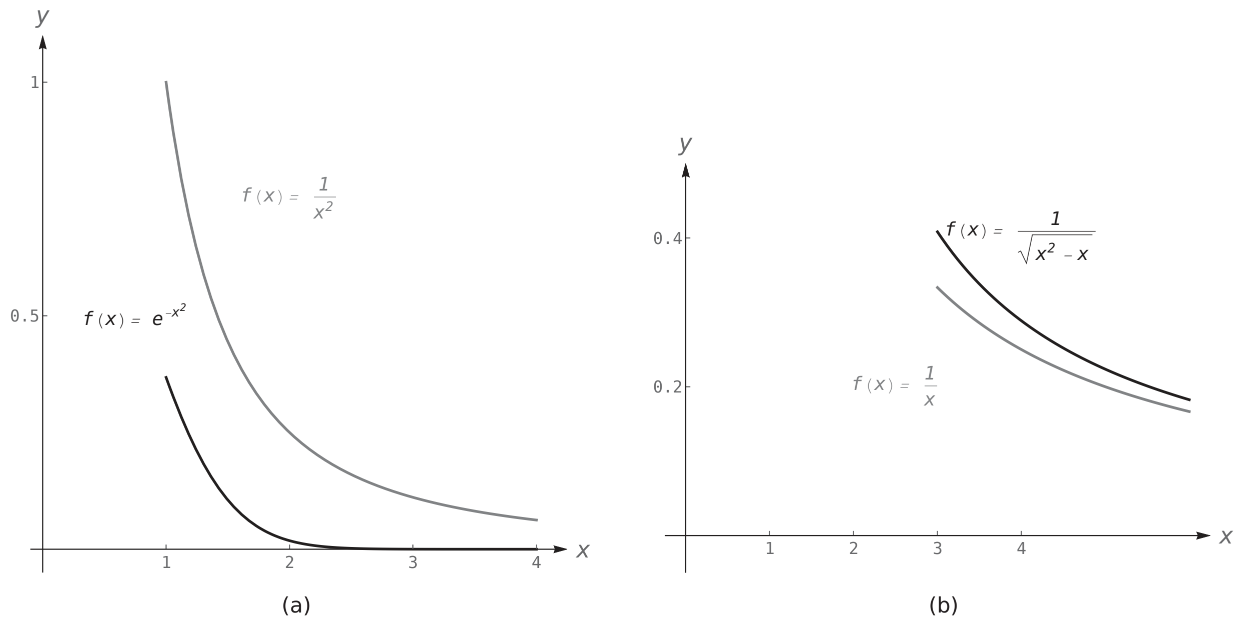


Figure 12.19: Graphs of $f(x) = e^{-x^2}$ and $f(x) = 1/x^2$ (a) and of $f(x) = 1/\sqrt{x^2-x}$ and $f(x) = 1/x$ (b) in Example 12.31.

Being able to compare unknown integrals to known integrals is very useful in determining convergence. However, some of our examples were a little too nice. For instance, it was convenient that $\frac{1}{x} < \frac{1}{\sqrt{x^2-x}}$, but what if the $-x$ were replaced with a $+2x+5$? That is, what can we say about the convergence of

$$\int_3^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\,dx?$$

We have

$$\frac{1}{x} > \frac{1}{\sqrt{x^2+2x+5}},$$

so we cannot use Theorem 12.9.

In cases like this (and many more) it is useful to employ the following theorem.

Theorem 12.12 (Limit comparison test for improper integrals)
Let f and g be continuous functions on $[a,+\infty[$ where $f(x)>0$ and $g(x)>0$ for all x. If

$$\lim_{x\to+\infty}\frac{f(x)}{g(x)} = L, \qquad 0 < L < +\infty,$$

then

$$\int_a^{+\infty} f(x)\,dx \text{ is convergent } \quad\Leftrightarrow\quad \int_a^{+\infty} g(x)\,dx \text{ is convergent },$$

and equivalently,

$$\int_a^{+\infty} f(x)\,dx \text{ is divergent } \quad\Leftrightarrow\quad \int_a^{+\infty} g(x)\,dx \text{ is divergent }.$$

We assume that L exists and is a positive finite number, and that the limit from a to $+\infty$ of g converges; we will show that the limit from a to $+\infty$ of f converges as well.

The definition of the limit tells us that, given the number $\epsilon = L/2$, there exists some M such that

$$\frac{L}{2} = L-\epsilon < \frac{f(x)}{g(x)} < L+\epsilon = \frac{3L}{2}$$

whenever $x > M$. So, for those values of x, we have that

$$\frac{L}{2}g(x) < f(x) < \frac{3L}{2}g(x)\,. \tag{12.27}$$

Let us now break the following integral in question into two pieces:

$$\int_a^{+\infty} f(x)\,dx = \int_a^M f(x)\,dx + \int_M^{+\infty} f(x)\,dx.$$

The first integral is of a continuous function on a closed, bounded interval, so we know that is finite. The convergence of the second integral is concluded by the following, which we can do because of Inequality (12.27):

$$\int_M^{+\infty} f(x)\,dx < \frac{3L}{2}\int_M^{+\infty} g(x)\,dx.$$

the last integral in this equation is given to converge (our assumption); therefore, by Theorem 12.9, the integral on the left converges as well. Hence, we conclude, as desired, that the integral of f converges.

Proving the other direction can be done similarly, or simply by observing that if $\lim\limits_{x\to+\infty} \frac{f(x)}{g(x)} = k$ exists and is positive, then $\lim\limits_{x\to+\infty} \frac{g(x)}{f(x)} = \frac{1}{k}$ must also exist and be positive.

The limit comparison test can as well be given for unproper integrals with an infinite range.

Theorem 12.13 (Limit comparison test for improper integrals with infinite range)
Let f and g be continuous functions on $[a, x_0[$ where $f(x) > 0$ and $g(x) > 0$ for all x. If

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = L, \qquad 0 < L < \infty,$$

then

$$\int_a^{x_0} f(x)\,dx \text{ is convergent } \Leftrightarrow \int_a^{x_0} g(x)\,dx \text{ is convergent,}$$

and equivalently,

$$\int_a^{x_0} f(x)\,dx \text{ is divergent } \Leftrightarrow \int_a^{x_0} g(x)\,dx \text{ is divergent.}$$

Example 12.32

Determine the convergence of

$$\int_3^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\,dx.$$

Solution

As x gets large, the denominator of the integrand will begin to behave much like $y = x$. So we compare

$$\frac{1}{\sqrt{x^2+2x+5}}$$

to $1/x$ using Theorem 12.12:

$$\lim_{x\to+\infty} \frac{1/\sqrt{x^2+2x+5}}{1/x} = \lim_{x\to+\infty} \frac{x}{\sqrt{x^2+2x+5}}.$$

The immediate evaluation of this limit returns ∞/∞, an indeterminate form. Using l'Hôpital's rule seems appropriate, but in this situation, it does not lead to useful results.

The trouble is the square root function. To get rid of it, we employ the following fact: If $\lim\limits_{x\to c} f(x) = L$, then $\lim\limits_{x\to c} f(x)^2 = L^2$. So we consider now the limit

$$\lim_{x\to+\infty} \frac{x^2}{x^2+2x+5}.$$

This converges to 1, meaning the original limit also converged to 1. As x gets very large, the function $\dfrac{1}{\sqrt{x^2+2x+5}}$ looks very much like $\dfrac{1}{x}$. Since we know that

$$\int_3^{+\infty} \frac{1}{x}\, dx$$

diverges, by Theorem 12.12 we know that

$$\int_3^{+\infty} \frac{1}{\sqrt{x^2+2x+5}}\, dx$$

also diverges. Figure 12.20 graphs $f(x) = 1/\sqrt{x^2+2x+5}$ and $f(x) = 1/x$, illustrating that as x gets large, the functions become indistinguishable.

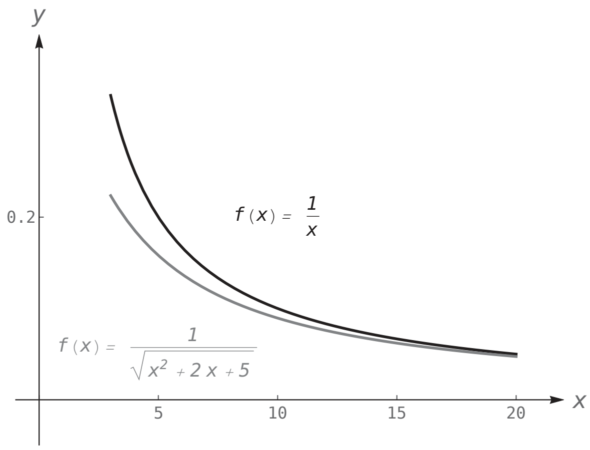


Figure 12.20: Graphing $f(x) = \frac{1}{\sqrt{x^2+2x+5}}$ and $f(x) = \frac{1}{x}$ in Example 12.32.

This chapter has explored many integration techniques. All of them effectively have one goal in common: rewrite the integrand in a new way so that the integration step is easier to see and implement. As stated before, integration is, in general, hard. It is easy to write a function whose antiderivative is impossible to write in terms of elementary functions, and even when a function does have an antiderivative expressible by elementary functions, it may be really hard to discover what it is. Mathematica, for instance, has approximately 1,000 pages of code dedicated to integration. Do not let this difficulty discourage you. There is great value in learning integration techniques, as they allow one to manipulate an integral in ways that can illuminate a concept for greater understanding. There is also great value in understanding the need for good numerical techniques

The next chapter stresses the uses of integration. We generally do not find antiderivatives for antiderivative's sake, but rather because they provide the solution to some type of problem. The following chapter introduces us to a number of different problems whose solution is provided by integration.

12.6 Exercises

12.6.1 Black-board exercises

1. Beschouw een partitie van het gegeven interval $[a,b]$ in n deelintervallen met gelijke breedte $\Delta x_i = (b-a)/n$. Bepaal de boven- en onder Riemann som voor de gegeven functies en gegeven waarde van n.

   (a) $f(x) = x, \quad [0,2], \quad n = 8$

   (b) $f(x) = \ln(x), \quad [1,2], \quad n = 5$

   (c) $f(x) = \sin(x), \quad [0,\pi], \quad n = 6$

   (d) $f(x) = \cos(x), \quad [0,2\pi], \quad n = 4$

   (e) $f(x) = x^2, \quad [-3,3], \quad n = 6$

   (f) $f(x) = \dfrac{1}{x}, \quad [1,9], \quad n = 4$

2. Druk de gegeven limiet uit als een bepaalde integraal.

   (a) $\displaystyle\lim_{n\to+\infty}\sum_{i=1}^{n}\frac{1}{n}\sqrt{\frac{i}{n}}$

   (b) $\displaystyle\lim_{n\to+\infty}\sum_{i=1}^{n}\frac{1}{n}\sqrt{\frac{i-1}{n}}$

   (c) $\displaystyle\lim_{n\to+\infty}\sum_{i=1}^{n}\frac{2}{n}\ln\left(1+\frac{2i}{n}\right)$

3. Als $a < b$ en f is continu in $[a,b]$, toon dan aan dat

$$\int_a^b \left(f(x)-\bar{f}\right) dx = 0,$$

   met $\bar{f}$ de gemiddelde waarde van f.

4. Bepaal de oppervlakte tussen de x-as en de curve beschreven door

$$f(x) = \begin{cases} x, & \text{als} \quad 0 < x \leq 1\,, \\ -2x+3, & \text{als} \quad 1 < x \leq 2\,, \\ -1, & \text{als} \quad 2 < x \leq 3\,, \\ 0, & \text{als} \quad x \leq 0 \vee x > 3\,. \end{cases}$$

5. Toon aan dat

$$I_n = \int_0^{\frac{\pi}{2}} \sin^n(x)\,dx = \left(\frac{n-1}{n}\right) I_{n-2}.$$

   Bereken $\displaystyle\int_0^{\frac{\pi}{2}} \sin^9(x)\,dx$.

6. Bepaal de afgeleide van de onderstaande functies.

(a) $F(x) = \int_{2}^{x^3+x} \frac{1}{t}\,dt$

(b) $F(x) = \int_{x^3}^{0} t^3\,dt$

(c) $F(t) = \int_{-\pi}^{t} \frac{\cos(y)}{1+y^2}\,dy$

(d) $F(t) = \int_{t}^{3} \frac{\sin(x)}{x}\,dx$

(e) $F(x) = x^2 \int_{0}^{x^2} \frac{\sin(u)}{u}\,du$

(f) $F(\theta) = \int_{\sin(\theta)}^{\cos(\theta)} \frac{dx}{1-x^2}$

(g) $F(x) = 3x \int_{4}^{x^2} e^{-\sqrt{t}}\,dt$

(h) $F(x) = \int_{x}^{x^2} (t+2)\,dt$

(i) $F(x) = \int_{\ln(x)}^{e^x} \sin(t)\,dt$

7. De zogenaamde error-functie wordt onder meer gebruikt voor het beschrijven van grondwaterstroming en wordt gegeven door

$$f(x) = \int_{0}^{x} e^{-u^2}\,du\,.$$

Echter, de integraal kan niet analytisch berekend worden en ook numerieke integratie is hier geen evidentie. De functie kan evenwel benaderd worden met behulp van een Taylor-reeksontwikkeling. Bepaal daartoe de MacLaurin-reeksontwikkeling van deze functie tot termen van de vierde orde.

8. Bepaal de onderstaande integralen.

(a) $\int_{4}^{9} \frac{1-\sqrt{x}}{1+\sqrt{x}}\,dx$

(b) $\int_{0}^{2} \sqrt{4-x^2}\,\frac{|x-1|}{x-1}\,dx$

9. Bepaal de onderstaande integralen.

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

(b) $\int \frac{2x+1}{4x^2+4x+3}\,dx$

(c) $\int \frac{\sin(x)}{\cos^6(x)}\,dx$

(d) $\int \cos^5(x)\,dx$

(e) $\int \frac{\sin(x)-\cos(x)}{\sin(x)+\cos(x)}\,dx$

(f) $\int \frac{dx}{\cos^2(x)\,\sqrt{1-4\tan^2(x)}}$

(g) $\int \frac{dx}{\big(\cos(x)+\sin(x)\big)^2}$

(h) $\int \ln\left(x+\sqrt{x^2+5}\right)dx$

(i) $\int \frac{2x-1}{x^2+x-6}\,dx$

(j) $\int \left(\frac{x-1}{x^2-5x+6}\right)^2 dx$

(k) $\int \frac{x^2+1}{x^2+2x+2}\,dx$

(l) $\int \frac{x+1}{(x^2+1)^{3/2}}\,dx$

(m) $\int \frac{dx}{\sqrt{4x-x^2}}$

(n) $\int e^{2x}\sin(4x)\,dx$

(o) $\displaystyle\int \sin^4(x)\cos^2(x)\,dx$

(p) $\displaystyle\int \frac{\cos(x)}{2\cos^2(x)+\sin(x)-1}\,dx$

(q) $\displaystyle\int \frac{dx}{\sin^2(x)\cos^4(x)}$

(r) $\displaystyle\int \frac{dx}{\sinh(x)}$

(s) $\displaystyle\int \tanh^3(x)\,dx$

10. Bepaal de onderstaande integralen.

(a) $\displaystyle\int \sin(x)\ \sinh(x)\,dx$

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

(c) $\displaystyle\int \frac{x^4}{x^3-8}\,dx$

(d) $\displaystyle\int \frac{dx}{x^4\sqrt{x^2-1}}$

(e) $\displaystyle\int \frac{3-4x}{\left(1-2\sqrt{x}\right)^2}\,dx$

(f) $\displaystyle\int \frac{dx}{\left(\tan(x)+1\right)\sin^2(x)}$

(g) $\displaystyle\int x\,e^{2x}\,dx$

(h) $\displaystyle\int \frac{5x}{\sqrt{x^4+1}}\,dx$

(i) $\displaystyle\int \sin\left(\frac{\pi}{4}-x\right)\sin\left(\frac{\pi}{4}+x\right)dx$

(j) $\displaystyle\int \frac{dx}{\sqrt[4]{5-x}+\sqrt{5-x}}$

(k) $\displaystyle\int x^2\ln\left(\sqrt{1-x}\right)dx$

(l) $\displaystyle\int \frac{2x-1}{2x+3}\,dx$

(m) $\displaystyle\int \sin(2x)\cos(2x)\,dx$

(n) $\displaystyle\int \frac{dx}{e^x+1}$

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

(p) $\displaystyle\int \frac{2x+3}{\left(x^2+x+1\right)^2}\,dx$

(q) $\displaystyle\int \sqrt{\frac{a+x}{a-x}}\,dx$

(r) $\displaystyle\int \frac{x-2\sqrt{x-1}}{1+\sqrt[4]{x-1}}\,dx$

(s) $\displaystyle\int \arctan(\sqrt{x})\,dx$

(t) $\displaystyle\int x^5\left(1+x^3\right)^{1/2}\,dx$

(u) $\displaystyle\int \frac{dx}{\sin^3(x)\ \cos^5(x)}$

(v) $\displaystyle\int \frac{dx}{\sin^6(x)}$

(w) $\displaystyle\int \frac{dx}{\sqrt{1+e^x}}$

(x) $\displaystyle\int 2^x\cosh(x)\,dx$

(y) $\displaystyle\int \sin^4(x)\,dx$

(z) $\displaystyle\int \frac{dx}{1+\cos(x)+\sin(x)}$

11. Onderzoek de convergentie van de onderstaande oneigenlijke integralen. Geef ook een verklaring.

(a) $\displaystyle\int_0^{\frac{\pi}{2}} \sec(x)\,dx$

(b) $\displaystyle\int_0^{+\infty} e^{-x}\sin(x)\,dx$

(c) $\displaystyle\int_{-1}^{8} x^{-\frac{2}{3}}\,dx$

(d) $\displaystyle\int_{-2}^{1} \frac{1}{\sqrt{|x|}}\,dx$

(e) $\displaystyle\int_0^{+\infty} xe^{x^2}\,dx$

(f) $\displaystyle\int_{-1}^{1} \frac{1}{x^4}\,dx$

(g) $\displaystyle\int_{0}^{e^2} \big(1+\ln(x)\big)\,dx$

(h) $\displaystyle\int_{0}^{+\infty} \frac{x^2}{x^5+1}\,dx$

(i) $\displaystyle\int_{0}^{+\infty} \frac{dx}{1+\sqrt{x}}$

(j) $\displaystyle\int_{0}^{+\infty} \frac{dx}{\sqrt{x}+x^2}$

(k) $\displaystyle\int_{-1}^{1} \frac{e^x}{x+1}\,dx$

(l) $\displaystyle\int_{2}^{+\infty} \frac{x\sqrt{x}}{x^2-1}\,dx$

(m) $\displaystyle\int_{0}^{\frac{\pi}{2}} \frac{\cos(x)}{\sqrt{x}}\,dx$

(n) $\displaystyle\int_{1}^{+\infty} \frac{\sin(x)}{x^2}\,dx$

12. Bereken de onderstaande integralen.

(a) $\displaystyle\int_{0}^{+\infty} \frac{e^{-ax}\sin(x)}{x}\,dx$

(b) $\displaystyle\int_{0}^{1} \frac{x^a-1}{\ln(x)}\,dx$

13. De Gamma-functie $\Gamma(x)$ wordt gedefinieerd door de oneigenlijke integraal

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}\,e^{-t}\,dt.$$

(a) Toon aan dat de integraal convergeert voor $x > 0$.

(b) Toon met behulp van partiële integratie aan dat voor $x > 0$ geldt dat $\Gamma(x+1) = x\Gamma(x)$.

(c) Toon aan dat $\Gamma(n+1) = n!$ voor $n = 0, 1, 2, \ldots$

(d) Veronderstel dat

$$\int_0^{+\infty} e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}.$$

Toon aan dat

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \qquad \text{en} \qquad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

14. Beschouw de midpointmethode over het interval $[a,b]$, dat we opdelen in n deelintervallen als

$$a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b\,,$$

waarbij $\Delta x = x_{i+1} - x_i = \frac{b-a}{n}$ voor alle $i = 1, \ldots, n+1$. Met behulp van deze methode benaderen we

$$S = \int_a^b f(x)dx$$

als

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f(m_i)\,,$$

met $m_i = \frac{x_i + x_{i+1}}{2}$ en is de totale fout op de benadering van S niets anders dan

$$E = \left| S - \hat{\boldsymbol{S}} \right| .$$

In wat volgt zullen we bewijzen dat het volgende geldt voor de bovengrens op de totale fout E van de midpointbenadering:

$$E \leq \frac{B(b-a)^3}{24n^2} ,$$

met B een constante.

Vul aan in deze bundel waar aangegeven door middel van een stippellijn.

Deze totale fout kan niet groter zijn dan de som der benaderingsfouten E_i voor de deelintervallen, dus er geldt dat

$$E \leq \sum_{i=1}^{n} E_i .$$

De lokale fout E_i is niets anders dan de netto-oppervlakte tussen de raaklijn aan f in $x = m_i$ en de grafiek van f over $[x_i, x_{i+1}]$. Zij $l(x)$ de lineaire benadering van f in $x = m_i$, d.i.

$$l(x) = \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

dan is

$$E_i = \left| \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \right| ,$$

waaruit volgt dat

$$E_i \leq \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \tag{12.28}$$

In wezen is de lineaire functie $l(x)$ niets anders dan de eerste-orde Taylor-veelterm in $x = m_i$ van f, waarvoor we weten dat de restterm gegeven wordt door

....

voor $z \in [x_i, x_{i+1}]$. Definiëren we nu de bovengrens op $\left| f''(z) \right|$ voor $z \in [x_i, x_{i+1}]$ als B, dan vinden we direct dat

$$\left| f(x) - l(x) \right| \leq \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

Bijgevolg vinden we als bovengrens voor het rechterlid van de ongelijkheid in Vergelijking (12.28)

.... (12.29)

Of meer expliciet na het uitrekenen van de voorkomende integraal in Vergelijking (12.29) en in acht nemend dat $x_{i+1} - m_i = \frac{\Delta x}{2}$ en $x_i - m_i = -\frac{\Delta x}{2}$

....

....

Hiervan gebruikmakend in Vergelijking (12.29), bekomen we als bovengrens voor E_i

....

Hieruit volgt voor de totale fout van de midpointbenadering E dat

....

12.6.2 Numerical integration

Het integreren van een functie is verre van evident. Hoewel in Secties 12.4 en 12.5 een arsenaal aan integratietechnieken aangereikt werd, moeten we vaststellen dat deze technieken in veel gevallen niet bruikbaar zijn. Doordat bijvoorbeeld de gezochte integraal gewoonweg niet als elementaire functie(s) kan worden uitgedrukt. Nog lastiger wordt het als we zelfs niet beschikken over het functievoorschrift van het integrandum, iets wat in de praktijk voortdurend voorkomt. Wat doen we in deze gevallen? We benaderen de (bepaalde) integraal als een som van eenvoudige, berekenbare oppervlaktes (Sectie 12.2.1). Deze manier van werken is conceptueel erg eenvoudig, maar is langdradig wanneer we een benadering met een aanvaardbare accuraatheid willen bekomen. Daarom zullen we hier enkele numerieke integratiemethoden implementeren en bestuderen in Matlab.

12.6.2.1 De midpointmethode

In Voorbeeld 12.4 hebben we een bepaalde integraal benaderd door de oppervlaktes van een reeks rechthoeken te sommeren. Voor de bepaalde integraal

$$S = \int_a^b f(x)\, dx\,,$$

bekwamen we deze reeks van n rechthoeken als volgt:

- deel het integratie-interval $[a, b]$ op in een partitie $\{x_1, x_2, ..., x_n, x_{n+1}\}$, waarbij
$$a = x_1 < x_2 < \cdots < x_n < x_{n+1} = b\,;$$
- de lengte van het i-de subinterval $[x_i, x_{i+1}]$ is de breedte van de i-de rechthoek;
- de hoogte van de i-de rechthoek wordt bepaald aan de hand van de linker-, rechter-, of midpointregel.

Indien voor een partitie gekozen wordt waarbij de breedte van de subintervallen constant is (stel Δx) en we de midpointregel toepassen, noemen we deze werkwijze de **midpointmethode**. De benadering $\hat{\boldsymbol{S}}$ van een integraal S, wordt dan als volgt berekend:

$$\begin{aligned} S = \int_a^b f(x)\, dx &\approx \Delta x f\left(\frac{x_1 + x_2}{2}\right) + \Delta x f\left(\frac{x_2 + x_3}{2}\right) + \cdots + \Delta x f\left(\frac{x_n + x_{n+1}}{2}\right) \\ &\approx \Delta x \sum_{i=1}^{n} f\left(\frac{x_i + x_{i+1}}{2}\right) = \hat{\boldsymbol{S}}\,. \end{aligned} \tag{12.30}$$

Als we Vergelijking (12.30) herschrijven als

$$\hat{\boldsymbol{S}} = \Delta x \sum_{i=1}^{n} f\,(m_i) = \Delta\, x \sum_{i=1}^{n} f\left(a + \frac{\Delta x}{2} + (i-1)\,\Delta x\right),$$

met $m_i = \frac{x_i + x_{i+1}}{2}$, d.i. het midden van interval $[x_i, x_{i+1}]$, kunnen we de midpointmethode als volgt vertalen naar uitvoerbare Matlab-code.

```
function [Sh, oppervlakten] = midpoint(f, interval, n)
%{
Midpointmethode voor de benadering van de integraal
van f over een gegeven interval [a,b]
Inputs:
        - f: integrandum (function handle)
        - interval: integratie-interval [a,b], opgegeven als een 1x2 rijvector
        - n: aantal deelintervallen

Outputs:
        - Sh: de benaderde integraal over [a,b] m.b.v. de midpointmethode
        - oppervlakten: oppervlakte van de rechthoeken
%}

% haal de begin- en eindpunten a en b uit het interval
```

```
a = interval(1);
b = interval(2);

% bereken de intervalbreedte deltax
deltax = (b - a)/n;

% bereken de middens van de deelintervallen
m = (a+deltax/2):deltax:b;   % m1 = a+deltax/2, m2 = a + deltax/2 + deltax, ...

% bereken de oppervlakte van de rechthoeken
oppervlakten = (deltax*f(m))';  % transponeren --> kolomvector

% bereken de benaderde integraal Sh
benadering = sum(oppervlakten);
end
```

De functie **`plotNumeriekeIntegratie(methode, f, interval, n)`** maakt een statische plot van f en de benadering van de integraal over het interval $[a,b]$ met een opgegeven numerieke integratiemethode. De inputs van deze functie worden als volgt gedefinieerd:

- methode: numerieke integratiemethode (`'midpoint'` of `'trapezium'`)
- f: integrandum
- interval: integratie-interval $[a,b]$
- n: aantal deelintervallen (defaultwaarde is 10)

Vraag 1.a Test de functie **`midpoint`** voor de bepaalde integraal uit Voorbeeld 12.4

$$S_1 = \int_{a_1}^{b_1} f_1(x)\,dx = \int_0^4 (4x - x^2)\,dx.$$

```
% anonieme functie die het integrandum van Voorbeeld 12.4 implementeert
f1 = @(x) ...

% oproep naar de functie midpoint
...
```

Vraag 1.b Maak met de functie **`plotNumeriekeIntegratie`** een plot van $f_1(x)$ en de benadering over het interval $[0,4]$ met $n = 10$.

```
% oproep naar de functie plotNumeriekeIntegratie
...
```

Je zou Figuur 12.21 moeten bekomen.

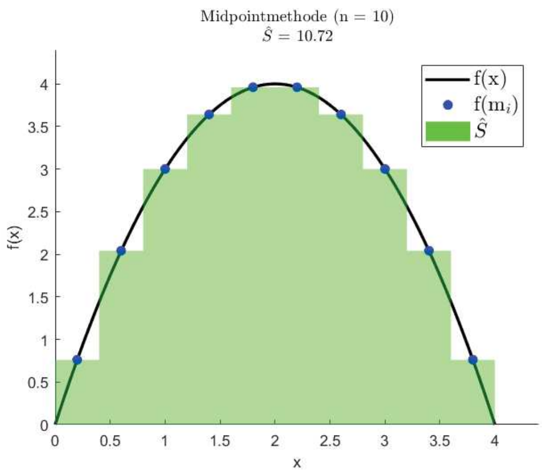


Figure 12.21: Benadering van de integraal van $f_1(x) = 4x - x^2$ over het interval $[0,4]$ met $n = 10$.

Vraag 1.c Ga met behulp van de onderstaande instructie(s) de rekentijd na voor $n = 100, 10^4, 10^6, \ldots$. Deze instructies zijn terug te vinden in het gegeven script `tijdsMeting.m`. Wat is de invloed van n?

```
f = @(x) 4*x - x.^2;  % te integreren functie
interval = [0, 4];  % integratie-interval
n = 100;            % aantal deelintervallen (aan te passen naar 10^4, 10^6, enz.)
k = 10;             % k metingen
tijd = zeros(k, 1);
for i = 1:k
    tic
    midpoint(f, interval, n);
    tijd(i) = toc;
end
gemTijd = mean(tijd)
```

12.6.2.2 Benaderingsfout

Om de accuraatheid van de numerieke integratie kwantitatief na te gaan, kunnen we gebruik maken van de relatieve benaderingsfout ϵ:

$$\epsilon = \frac{|S - \hat{\mathbf{S}}|}{|S|}.$$

Om het effect van het aantal deelintervallen n op ϵ na te gaan, maken we gebruik van de gegeven functie **`plotError(S, methode, f, interval, nBereik)`**. Deze plot de relatieve benaderingsfout ϵ van een opgegeven numerieke integratiemethode in functie van het aantal deelintervallen n. De inputs van deze functie worden als volgt gedefinieerd:

- `S`: exacte waarde van de integraal
- methode: numerieke integratiemethode (`'midpoint'` of `'trapezium'`)
- `f`: integrandum
- `interval`: integratie-interval $[a, b]$
- `nBereik`: bereik van n-waarden $[n_{min}, n_{max}]$ $(default = [1, 20])$

Vraag 2.a We willen de functie **`plotError`** testen voor de benadering van S_1 voor n gaande van 1 tot 20. Hiervoor hebben we S_1 nodig. Bereken daartoe eerst de exacte waarde S_1 (met de hand of met Mathematica) en geef deze hieronder in.

```
S1 =     % aan te vullen met de exacte waarde van de integraal van f1 over [0, 4]
```

Vraag 2.b Plot nu met de functie **`plotError`** de relatieve benaderingsfout ϵ van de integraal van $f_1(x)$ voor n gaande van 1 tot 20.

```
% aan te vullen met een oproep naar de functie plotError
...
```

Je zou Figuur 12.22 moeten bekomen.

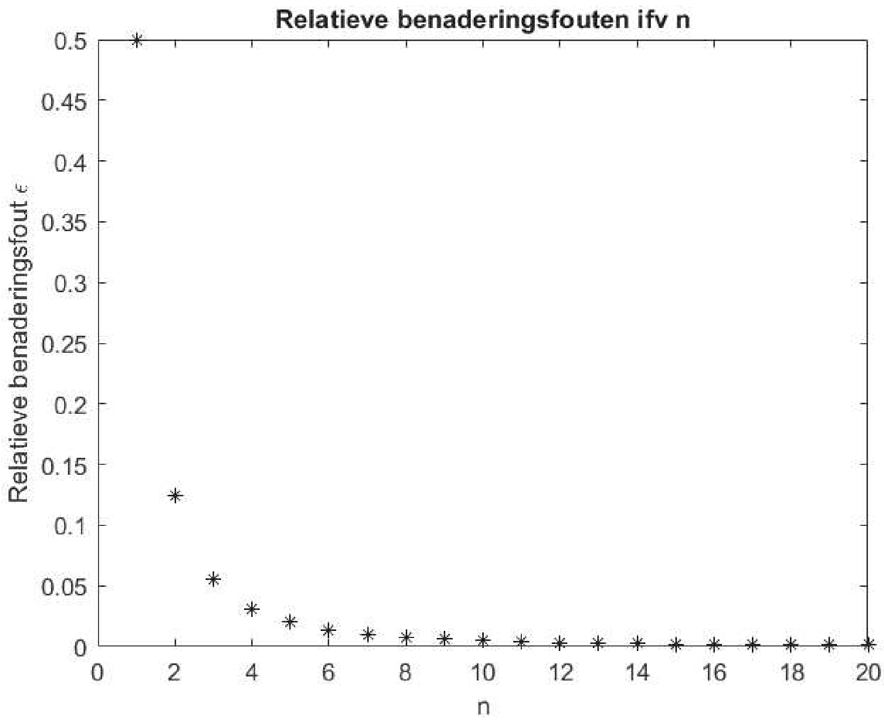


Figure 12.22: Relatieve benaderingsfouten voor de integraal van $f_1(x) = 4x - x^2$ met $n = 1, 2, \ldots, 20$.

Vraag 2.c Implementeer het integrandum van de volgende bepaalde integralen en bereken (indien mogelijk) de exacte waarde:

- $S_2 = \int_{a_2}^{b_2} f_2(x)\, dx = \int_0^2 \sin(2x)\cos(2x)\, dx\,,$
- $S_3 = \int_{a_3}^{b_3} f_3(x)\, dx = \int_{-\frac{\pi}{4}}^{\frac{\pi}{2}} \sin\left(x^3\right) dx\,.$

Wat merk je op?

```
% anonieme functie die f2 implementeert
f2 = @(x) = ... % aan te vullen
```

```
% exacte waarde van de integraal van f2 over [0, 2]
S2 = ... % aan te vullen


% anonieme functie die f3 implementeert
f3 = @(x) = ... % aan te vullen

% exacte waarde van de integraal van f3 over [-pi/4, pi/2] (indien uit te rekenen)
S3 = ... % aan te vullen
```

Vraag 2.d Bereken voor de integraal van $f_2(x)$ en $f_3(x)$ de benadering met de midpointmethode, en plot het verloop van de relatieve benaderingsfout in functie van n (voor n gaande van 1 tot 20) met **`plotError`**.

```
% benadering van f2

% fouten voor f2

% benadering van f3

% fouten voor f3 (indien mogelijk)
```

12.6.2.3 De trapeziummethode

De midpointmethode benadert een bepaalde integraal als de som van rechthoeken. Hierbij wordt het integrandum in elk deelinterval $]x_i, x_{i+1}[$ benaderd door een constante functie $f(m_i)$. In deelintervallen waarover de functiewaarde sterk verandert, is deze benadering echter niet nauwkeurig. Dit valt goed te zien in de plot van f_3 in de vorige vraag.

Een alternatief voor de midpointmethode is de **trapeziummethode**, waarbij de oppervlaktes in de deelintervallen worden benaderd door - verrassing - trapezia. Deze methode benadert het integrandum over het deelinterval $[x_i, x_{i+1}]$ als een rechte gaande van $\big(x_i, f(x_i)\big)$ tot $\big(x_{i+1}, f(x_{i+1})\big)$. De bepaalde integraal wordt dan benaderd als volgt:

$$S = \int_a^b f(x)\,dx \quad \approx \quad \Delta x \sum_{i=1}^{n} \frac{f(x_i) + f(x_{i+1})}{2} = \hat{\boldsymbol{S}}\,. \tag{12.31}$$

Vraag 3.a Implementeer de trapeziummethode door de onderstaande functie aan te vullen op de plaatsen aangegeven door "...".

Tip 1: Alle functies en technieken nodig voor de implementatie worden aangereikt in de implementatie van de midpointmethode.
Tip 2: Er zijn meerdere manieren om de trapeziummethode te implementeren, de een al efficiënter dan de andere.

```
function [Sh, oppervlakten] = trapezium(f, interval, n)
%{
Trapeziummethode voor de benadering van de integraal
van f over een gegeven interval [a,b]
Inputs:
        - f: integrandum (function handle)
        - interval: integratie-interval [a,b], opgegeven als een 1x2 rijvector
```

```
        - n: aantal deelintervallen

Outputs:
        - Sh: de benaderde integraal over [a,b] m.b.v. de trapeziummethode
        - oppervlakten: oppervlakte van de trapezia
%}
% haal de begin- en eindpunten a en b uit het interval
...
...

% bereken de intervalbreedte deltax
...

% bereken de begin- en eindpunten van de deelintervallen
...

% bereken de oppervlakte van de trapezia
% dit kan met en zonder een for-lus
...
...
...
...
...

% bereken de benaderde integraal Sh
...
end
```

Vraag 3.b Vergelijk de resultaten van de midpoint- en trapeziummethode voor f_1, f_2 en f_3. Merk je bepaalde verschillen op? Welke methode verkies jij en waarom?

```
% benadering voor f1(x) met midpoint

% benadering voor f1(x) met trapezium

% benadering voor f2(x) met midpoint

% benadering voor f2(x) met trapezium

% benadering voor f3(x) met midpoint

% benadering voor f3(x) met trapezium
```

Vraag 3.c(*) Beschouw de volgende bepaalde integraal:

$$S_{tot} = \int_a^b f_t(x)\,dx = \int_a^b \big(f_s(x) + f_r(x)\big)\,dx.$$

Het integrandum $f_t(x)$ is hier een som van twee deelfuncties, nl. een signaalfunctie $f_s(x)$ en ruisfunctie $f_r(x)$, waarvan we de wiskundige vorm niet kennen. Dit komen we in de praktijk voortdurend tegen, bijvoorbeeld wanneer we een meettoestel gebruiken waar er zich (na een tijd) vuil op afzet. Hierbij is x dan de tijd, $f_s(x)$ de te meten grootheid i.f.v. de tijd en $f_r(x)$ de verstoring van het signaal door vuilafzetting i.f.v. de tijd.

Download de functies `fs`, `fr` en `ft` en ga voor beide numerieke methoden na wat het effect is van de ruis op de benaderde integraal van $f_t(x)$ over het interval $[0, 20]$. Is er een methode te verkiezen boven de andere?

As you will find in multivariable calculus, there is often a number of solutions for any given problem.

— John Nash —

16

Functions of several variables

A function of the form $y = f(x)$ is a function of a single variable; given a value of x, we can find a value y. Even the vector–valued functions of Chapter 15 are single–variable functions; the input is a single variable though the output is a vector.

There are many situations where a desired quantity is a function of two or more variables. For instance, wind chill is measured by knowing the temperature and wind speed; the volume of a gas can be computed knowing the pressure and temperature of the gas; to compute a baseball player's batting average, one needs to know the number of hits and the number of at–bats.

This chapter studies multivariable functions, that is, functions with more than one input.

16.1 Introduction to multivariable functions

16.1.1 Functions of two variables

We start with a definition of a function of two variables.

Definition 16.1 (Function of two variables)
Let D be a subset of $\mathbb{R}^2$. A **function f of two variables** (*functie van twee veranderlijken*) is a rule that assigns each pair (x, y) in D a value $z = f(x, y)$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

Example 16.1
Let

$$f(x, y) = \sqrt{1 - \frac{x^2}{9} - \frac{y^2}{4}}.$$

Find the domain and range of f.

Solution

The domain is all pairs (x,y) allowable as input in f. Because of the square root, we need (x,y) such that:

$$\begin{aligned} & 1-\frac{x^2}{9}-\frac{y^2}{4} \geq 0 \\ \Leftrightarrow \qquad & \frac{x^2}{9}+\frac{y^2}{4} \leq 1 \end{aligned}$$

The above equation describes an ellipse and its interior. We can represent the domain D in set notation as

$$D=\left\{(x,y) \;\middle|\; \frac{x^2}{9}+\frac{y^2}{4} \leq 1\right\}.$$

The range is the set of all possible output values. The square root ensures that all output is positive. Since the x and y terms are squared, then subtracted, inside the square root, the largest output value comes at $x=0$, $y=0$: $f(0,0)=1$. Thus the range R is the interval $[0,1]$.

Definition 16.2 (Graph of a function of two variables)
The **graph** of a function f of two variables is the set of all points $\big(x,y,f(x,y)\big)$ where (x,y) is in the domain of f. This creates a **surface** (*oppervlak*) in space.

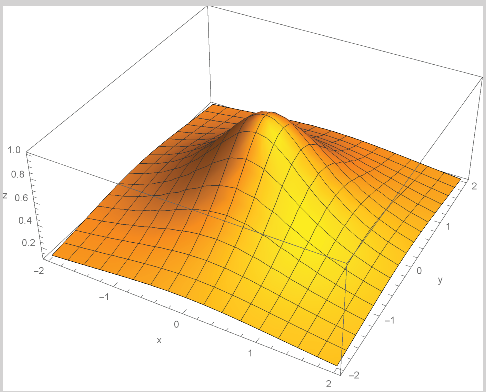
One can begin sketching a graph by plotting points, but this has limitations. Consider Figure 16.1(a) where 25 points have been plotted of

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

More points have been plotted than one would reasonably want to do by hand, yet it is not clear at all what the graph of the function looks like. Technology allows us to plot lots of points, connect adjacent points with lines and add shading to create a graph like Figure 16.1(b) which does a far better job of illustrating the behaviour of f. More specifically, in Mathematica, a function of two variables can be plotted using the command **Plot3D** as follows

```
In[25]:= Plot3D[{1/(x^2 + y^2 + 1)}, {x, -2, 2}, {y, -2, 2},  AxesLabel → {"x", "y", "z"}]
```

Out[25]=



Of course, many options are available to format such graphs according to one's preference. These can be checked in the Documentation Center.

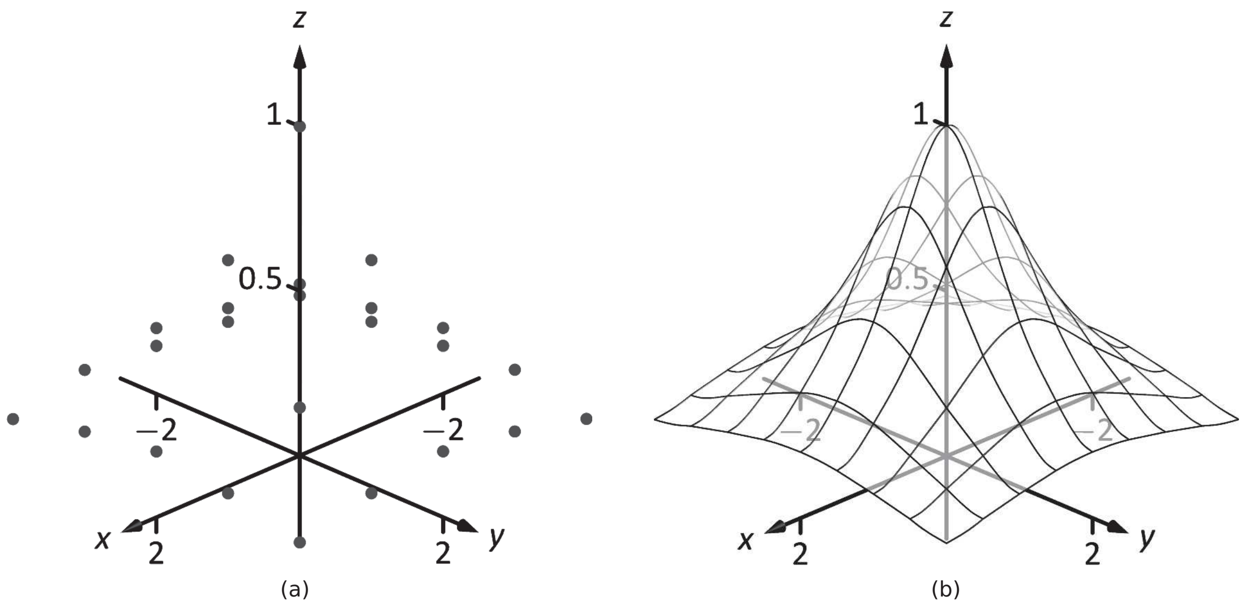


Figure 16.1: Graphing a function of two variables.

While technology is readily available to help us graph functions of two variables, there is still a paper–and–pencil approach that is useful to understand and master as it, combined with high–quality graphics, gives one great insight into the behaviour of a function. This technique is known as sketching **level curves** (*niveaukromme*).

It may be surprising to find that the problem of representing a three dimensional surface on paper is familiar to most people. Topographical maps, like the one of Dinant shown in Figure 16.2, represent the surface of Earth by indicating points with the same elevation with **contour lines** (*countourlijn*). The elevations marked are equally spaced; in this example, each thin line indicates an elevation change in 10m increments and each thick line indicates a change of 50m. When lines are drawn close together, elevation changes rapidly. When lines are far apart, elevation changes more gradually as one has to walk farther to rise 10m.

Given a function $z = f(x, y)$, we can draw a topographical map of f by drawing **level curves** (or, contour lines). A level curve at $z = c$ is a curve in the xy-plane such that for all points (x, y) on the curve, $f(x, y) = c$. When drawing level curves, it is important that the c-values are spaced equally apart as that gives the best insight to how quickly the elevation is changing.

Figure 16.2: The topographical map of Dinant displays elevation by drawing contour lines, along which the elevation is constant.

Example 16.2

Let

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

Find level curves.

Solution

We begin by setting $f(x,y) = c$ for an arbitrary c and seeing if algebraic manipulation of the equation reveals anything significant.

$$\frac{x+y}{x^2+y^2+1} = c \qquad \Leftrightarrow \qquad x^2 - \frac{1}{c}x + y^2 - \frac{1}{c}y = -1.$$

We recognize this as a circle, though the centre and radius are not yet clear. By completing the square, we obtain:

$$\left(x - \frac{1}{2c}\right)^2 + \left(y - \frac{1}{2c}\right)^2 = \frac{1}{2c^2} - 1$$

a circle centred at $\big(1/(2c), 1/(2c)\big)$ with radius $\sqrt{1/(2c^2)-1}$, where $|c| < 1/\sqrt{2}$. The level curves for $c = \pm 0.2$, ± 0.4 and ± 0.6 are sketched in Figure 16.3(a). To help illustrate elevation, we use dashed lines where $c < 0$. There is one special level curve, when $c = 0$. The level curve in this situation is $x + y = 0$, the line $y = -x$.

In Figure 16.3(b) we see a graph of the surface. Note how the y-axis is pointing away from the viewer to more closely resemble the orientation of the level curves in Figure 16.3(a). Seeing the level curves helps us understand the graph. For instance, the graph does not make it clear that one can walk along the line $y = -x$ without elevation change, though the level curve does.

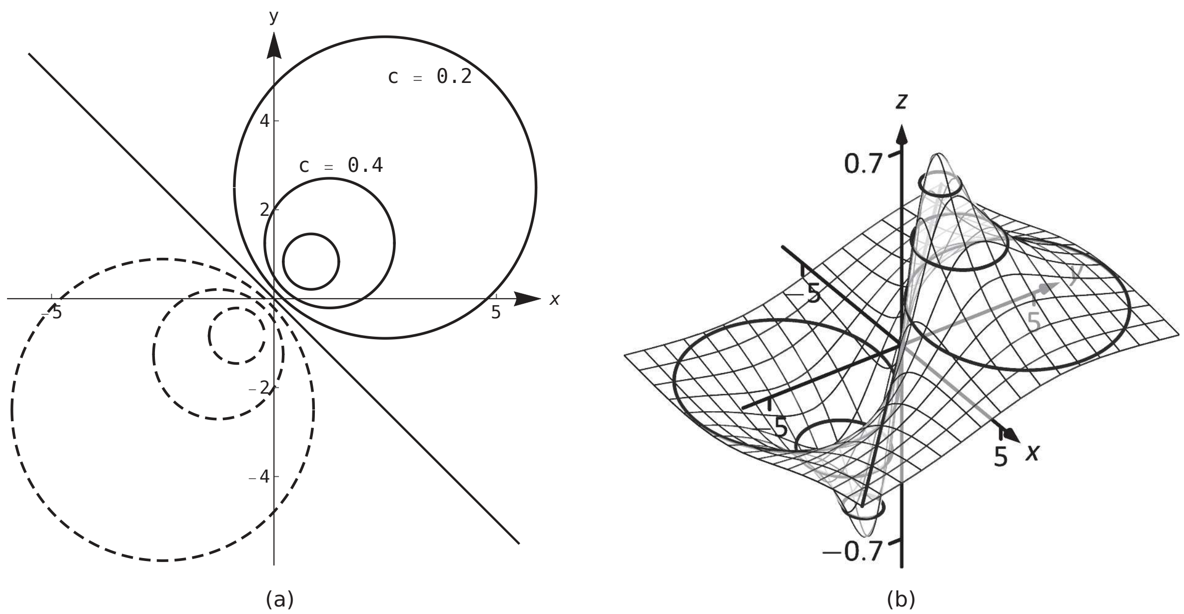


Figure 16.3: Graphing the level curves in Example 16.2.

16.1.2 Functions of *n* variables

We extend our study of multivariable functions to functions of n variables.

Definition 16.3 (Function of *n* variables)
Let D be a subset of $\mathbb{R}^n$. A **function** f **of** n **variables** is a rule that assigns each $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in D a value $w = f(x_1, x_2, \ldots, x_n) = f(\mathbf{x})$ in $\mathbb{R}$. D is the domain of f; the set of all outputs of f is the range.

Note that in this definition, we are using the notation $\mathbf{x}$ to abbreviate the n-tuple $(x_1, x_2, \ldots, x_n)$. It is very difficult to produce a meaningful graph of a function of three variables. A function of one variable is a curve drawn in 2 dimensions; a function of two variables is a surface drawn in 3 dimensions; a function of n variables is a **hypersurface** (*hyperoppervlak*) drawn in $n+1$ dimensions.

There are a few techniques one can employ to try to picture a graph of three variables. One is an analogue of level curves: **level surfaces** (*niveau-oppervlak*). Given $w = f(x, y, z)$, the level surface at $w = c$ is the surface in space formed by all points (x, y, z) where $f(x, y, z) = c$.

Example 16.3

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0, 0, 0)$,

$$I(x, y, z) = \frac{k}{x^2 + y^2 + z^2}$$

for some constant k.

Let $k = 1$; find the level surfaces of I.

Solution

We can answer this question using common sense. If energy is emanating from the origin, its intensity will be the same at all points equidistant from the origin. That is, at any point on the surface of a sphere centred at the origin, the intensity should be the same. Therefore, the level surfaces are spheres. We now find this mathematically. The level surface at $I = c$ for $c > 0$ is defined by

$$c = \frac{1}{x^2 + y^2 + z^2}.$$

A small amount of algebra reveals

$$x^2 + y^2 + z^2 = \frac{1}{c}.$$

Given an intensity c, the level surface $I = c$ is a sphere of radius $1/\sqrt{c}$, centred at the origin. Table 16.1 gives the radii of the spheres for given c-values. Normally one would use equally spaced c-values, but these values have been chosen purposefully. At a distance of 0.25 from the point source, the intensity is 16; to move to a point of half that intensity, one just moves out 0.1 to 0.35 – not much at all. To again halve the intensity, one moves 0.15, a little more than before.

Table 16.1: A table of c-values and corresponding radius r of the spheres of constant value in Example 16.3.

c	16.	8.	4.	2.	1.	0.5	0.25	0.125	0.0625
r	0.25	0.35	0.5	0.71	1	1.41	2	2.83	4

Note how each time the intensity is halved, the distance required to move away grows. We conclude that the closer one is to the source, the more rapidly the intensity changes.

16.2 Limits and continuity of multivariable functions

This section investigates what it means for multivariable functions to be continuous.

16.2.1 Introductory concepts and definitions

We begin with a series of definitions. We are used to open and closed intervals. We need analogous definitions for open and closed sets in the n-dimensional space.

Definition 16.4 (Open ball, boundary and interior points, open, closed and bounded sets)
An **open ball** (*open bal*) B in $\mathbb{R}^n$ centred at $\mathbf{x}_0$ with radius r is the set of all points $\mathbf{x}$ such that $d(\mathbf{x}, \mathbf{x}_0) < r$.

Let S be a set of points in $\mathbb{R}^n$. A point P in $\mathbb{R}^n$ is a **boundary point** (*randpunt*) of S if all open disks centred at P contain both points in S and points not in S.

A point P in S is an **interior point** (*inwendig punt*) of S if there is an open disk centred at P that contains only points in S.

A set S is **open** (*open*) if every point in S is an interior point.

A set S is **closed** (*gesloten*) if it contains all of its boundary points.

A set S is **bounded** (*begrensd*) if there is an $M > 0$ such that the open disk, centred at the origin with radius M, contains S. A set that is not bounded is unbounded.

Figure 16.4 shows several sets in the xy-plane. In each set, point P_1 lies on the boundary of the set as all open disks centred there contain both points in, and not in, the set. In contrast, point P_2 is an interior point for there is an open disk centred there that lies entirely within the set. The set depicted in Figure 16.4(a) is a closed set as it contains all of its boundary points. The set in Figure 16.4(b) is open, for all of its points are interior points (or, equivalently, it does not contain any of its boundary points). The set in Figure 16.4(c) is neither open nor closed as it contains some of its boundary points.

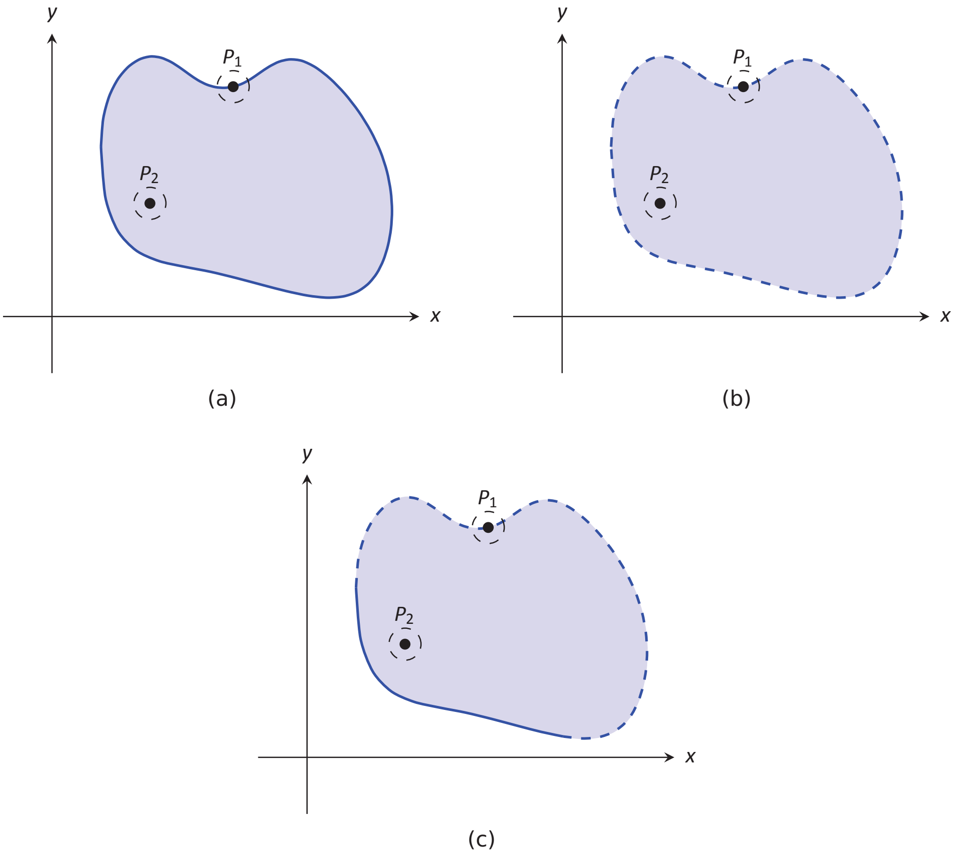


Figure 16.4: Illustrating open and closed sets in the xy-plane.

Example 16.4

Determine if the domain of the functions

1. $f(x,y)=\sqrt{1-\frac{x^2}{9}-\frac{y^2}{4}}$
2. $g(x,y)=\frac{1}{x-y}$

is open, closed, or neither, and if it is bounded.

Solution

1. The domain of this function was found in Example 16.1 to be

$$D=\left\{(x,y)\;\middle|\;\frac{x^2}{9}+\frac{y^2}{4}\leq 1\right\},$$

the region bounded by the ellipse $\frac{x^2}{9}+\frac{y^2}{4}=1$. Since the region includes the boundary, the set contains all of its boundary points and hence is closed. The region is bounded as a disk of radius 4, centred at the origin, contains D.

2. As we cannot divide by 0, we find the domain to be $D=\{(x,y)\mid x-y\neq 0\}$. In other words, the domain is the set of all points (x,y) not on the line $y=x$. The domain is sketched in Figure 16.5. Note how we can draw an open disk around any point in the domain that lies

entirely inside the domain, and also note how the only boundary points of the domain are the points on the line $y = x$. We conclude the domain is an open set. The set is unbounded.

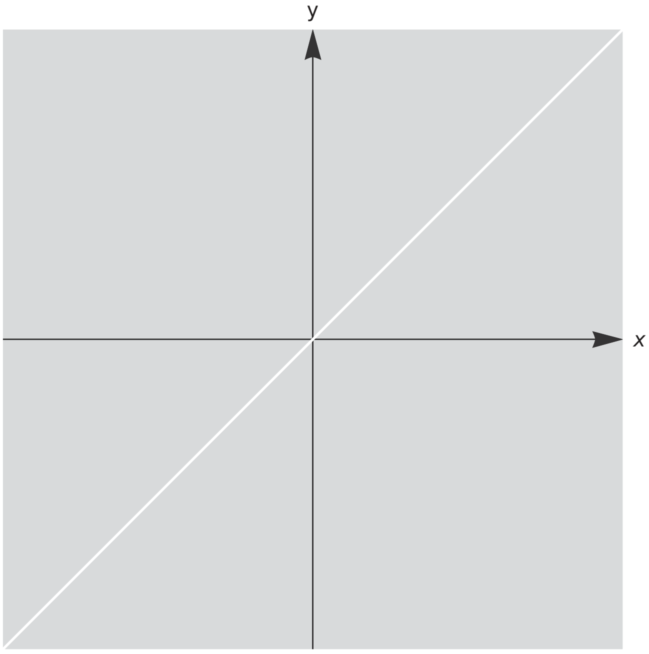


Figure 16.5: Sketching the domain of the function in Example 16.4.2.

16.2.2 Limits

We will say that

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$$

means if the point (x,y) is really close to the point (x_0,y_0), then $f(x,y)$ is really close to L. The formal definition for a function of n variables is given below.

Definition 16.5 (Limit of a function of *n* variables)
Let S be a set containing $P = \mathbf{x}_0$ where every open disk centred at P contains points in S other than P, i.e. P is a limit point, let f be a function of two variables defined on S, except possibly at P, and let L be a real number. The **limit of $f(\mathbf{x})$ as $\mathbf{x}$ approaches $\mathbf{x}_0$** is L, denoted

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = L,$$

means that given any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{x}$ in S, where $\mathbf{x} \neq \mathbf{x}_0$, if $\mathbf{x}$ is in the open ball centred at $\mathbf{x}_0$ with radius δ, then $|f(\mathbf{x}) - L| < \varepsilon$.

Note that we now define limits over a set S in the plane (where S does not have to be open). As planar sets can be far more complicated than intervals, our definition adds the restriction "*...* where every open disk centred at P contains points in S other than P." This means that P should be a so-called **limit point** (*ophopingspunt*) of the set S. This in contrast to a so-called **isolated point** (*gëisoleerd punt*) Q of S for which there exists a neighbourhood of Q which does not contain any other points of S.

The concept behind Definition 16.5 is sketched in Figure 16.6. Given $\varepsilon > 0$, find $\delta > 0$ such that if (x, y) is any point in the open disk centred at (x_0, y_0) in the xy-plane with radius δ, then $f(x, y)$ should be within ε of L.

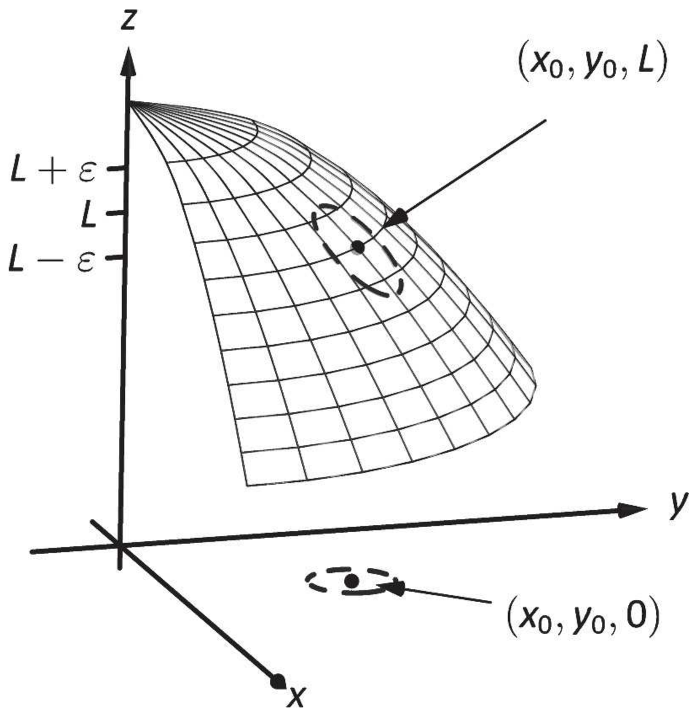


Figure 16.6: Illustrating the definition of a limit of a function of two variables.

Computing limits using this definition is rather cumbersome. The following properties allow us to evaluate limits much more easily. For that purpose, let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, let $\mathbf{x}_0$ have real components $(x_0)_i$ and let b, L and K be real numbers, let n be a positive integer, and let f and g be functions with the following limits:

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = L \quad \text{and} \quad \lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) = K.$$

The following limits hold.

- **Constants:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} b = b$$

- **Identity**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} x_i = (x_0)_i$$

- **Sums/Differences:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \big(f(\mathbf{x}) \pm g(\mathbf{x})\big) = L \pm K$$

- **Scalar Multiples:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} b \cdot f(\mathbf{x}) = bL$$

- **Products:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) \cdot g(\mathbf{x}) = LK$$

- **Quotients:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{f(\mathbf{x})}{g(\mathbf{x}} = \frac{L}{K}, \qquad (K \neq 0)$$

- **Powers:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0} [f(\mathbf{x})]^n = L^n$$

These properties, combined with the ones we introduced in Chapter 8, allow us to evaluate many limits. For instance, we can easily evaluate

$$\lim_{(x,y)\to(1,\pi)} \left(\frac{y}{x} + \cos(xy)\right) = \frac{\pi}{1} + \cos(\pi) = \pi - 1.$$

This limit may as well be evaluated in Mathematica with a nested application of the command `Limit`.

```
In[26]:= Limit[Limit[y/x + Cos[x*y], x → 1], y → Pi]

Out[26]= -1+π
```

When dealing with functions of a single variable we also considered one–sided limits and stated

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

if, and only if

$$\lim_{\substack{x\to c\\ >}} f(x) = L \quad \text{and} \quad \lim_{\substack{x\to c\\ <}} f(x) = L.$$

That is, the limit is L if and only if $f(x)$ approaches L when x approaches c from either direction, the left or the right.

In the plane, there are infinitely many directions from which (x,y) might approach (x_0,y_0). In fact, we do not have to restrict ourselves to approaching (x_0,y_0) from a particular direction, but rather we can approach that point along a path that is not a straight line. It is possible to arrive at different limiting values by approaching (x_0,y_0) along different paths. If this happens, we say that $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$ does not exist. This is analogous to the left and right hand limits of single variable functions not being equal.

Our theorems tell us that we can evaluate most limits quite simply, without worrying about paths. When indeterminate forms arise, the limit may or may not exist. If it does exist, it can be difficult to prove this as we need to show the same limiting value is obtained regardless of the path chosen. The case where the limit does not exist is often easier to deal with, for we can often pick two paths along which the limit is different.

Example 16.5

1. Show

$$\lim_{(x,y)\to(0,0)} \frac{3xy}{x^2+y^2}$$

does not exist by finding the limits along the lines $y = mx$.

2. Show

$$\lim_{(x,y)\to(0,0)} \frac{\sin(xy)}{x+y}$$

does not exist by finding the limit along the path $y = -\sin(x)$.

Solution

1. Evaluating this limit along the lines $y = mx$ means replace all y's with mx and evaluating the resulting limit:

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{3x(mx)}{x^2+(mx)^2} &= \lim_{x\to 0} \frac{3mx^2}{x^2(m^2+1)}\\ &= \lim_{x\to 0} \frac{3m}{m^2+1}\\ &= \frac{3m}{m^2+1}.\end{aligned}$$

While the limit exists for each choice of m, we get a different limit for each choice of m. That is, along different lines we get differing limiting values, meaning the limit does not exist.

2. We are to show that $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ does not exist by finding the limit along the path $y = -\sin(x)$. First, however, consider the limits found along the lines $y = mx$ as done above.

$$\begin{aligned}\lim_{(x,mx)\to(0,0)} \frac{\sin\big(x(mx)\big)}{x+mx} &= \lim_{x\to 0} \frac{\sin(mx^2)}{x(m+1)}\\ &= \lim_{x\to 0} \frac{\sin(mx^2)}{x}\cdot\frac{1}{m+1}.\end{aligned}$$

By applying L'Hôpital's rule, we can show this limit is 0 except when $m = -1$, that is, along the line $y = -x$. This line is not in the domain of f, so we have found the following fact: along every line $y = mx$ in the domain of f,

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

Now consider the limit along the path $y = -\sin(x)$:

$$\lim_{(x,-\sin(x))\to(0,0)} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)} = \lim_{x\to 0} \frac{\sin\big(-x\sin x\big)}{x-\sin(x)}.$$

Now apply L'Hôpital's rule twice:

$$\begin{aligned}&= \lim_{x\to 0} \frac{\cos\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))}{1-\cos(x)} \qquad \left(=\frac{0}{0}\right)\\ &= \lim_{x\to 0} \frac{-\sin\big(-x\sin(x)\big)(-\sin(x)-x\cos(x))^2+\cos\big(-x\sin(x)\big)(-2\cos(x)+x\sin(x))}{\sin(x)}\\ &= \frac{-2}{0}.\end{aligned}$$

It follows that the limit does not exist. Step back and consider what we have just discovered. Along any line $y = mx$ in the domain of the $f(x,y)$, the limit is 0. However, along the path $y = -\sin(x)$, which lies in the domain of $f(x,y)$ for all $x \neq 0$, the limit does not exist. Since the limit is not the same along every path to $(0,0)$, we say that the studied limit does not exist.

Example 16.6

Let

$$f(x,y) = \frac{5x^2y^2}{x^2+y^2}.$$

Find $\lim_{(x,y)\to(0,0)} f(x,y)$.

Solution

It is relatively easy to show that along any line $y = mx$, the limit is 0. This is not enough to prove that the limit exists, as demonstrated in the previous example, but it tells us that if the limit does exist then it must be 0.

To prove the limit is 0, we apply Definition 16.5. Let $\varepsilon > 0$ be given. We want to find $\delta > 0$ such that if $\sqrt{(x-0)^2+(y-0)^2} < \delta$, then $|f(x,y)-0| < \varepsilon$.

Set $\delta < \sqrt{\varepsilon/5}$. Note that $\left|\frac{5y^2}{x^2+y^2}\right| < 5$ for all $(x,y) \neq (0,0)$, and that if $\sqrt{x^2+y^2} < \delta$, then $x^2 < \delta^2$.

Let $\sqrt{(x-0)^2+(y-0)^2} = \sqrt{x^2+y^2} < \delta$. Consider $|f(x,y)-0|$:

$$\begin{aligned}
|f(x,y)-0| &= \left|\frac{5x^2y^2}{x^2+y^2} - 0\right| \\
&= \left|x^2 \cdot \frac{5y^2}{x^2+y^2}\right| \\
&< \delta^2 \cdot 5 \\
&< \frac{\varepsilon}{5} \cdot 5 \\
&= \varepsilon.
\end{aligned}$$

Thus if $\sqrt{(x-0)^2+(y-0)^2} < \delta$ then $|f(x,y)-0| < \varepsilon$, which is what we wanted to show. Thus

$$\lim_{(x,y)\to(0,0)} \frac{5x^2y^2}{x^2+y^2} = 0\,.$$

16.2.3 Continuity

Definition 8.3 defines what it means for a function of one variable to be continuous. In brief, it meant that the graph of the function did not have breaks, holes, jumps, etc. We define continuity for functions of two variables in a similar way as we did for functions of one variable.

Definition 16.6 (Continuity)
Let a function $f(\mathbf{x})$ be defined on a set S containing the point $\mathbf{x}_0$.

1. f is continuous at $\mathbf{x}_0$ if $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$.
2. f is **continuous on** S (*continu over*) if f is continuous at all points in S. If f is continuous at all points in $\mathbb{R}^n$, we say that f is continuous everywhere.

Example 16.7

Let

$$f(x,y)=\begin{cases}\dfrac{\cos(y)\sin(x)}{x}, & x\neq 0\\ \cos(y), & x=0.\end{cases}$$

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

Solution

To determine if f is continuous at $(0,0)$, we need to compare $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ to $f(0,0)$. Applying the definition of f, we see that $f(0,0)=\cos(0)=1$.

We now consider the limit $\lim\limits_{(x,y)\to(0,0)} f(x,y)$. Substituting 0 for x and y in $f(x,y)$ returns the indeterminate form "0/0", so we need to do more work to evaluate this limit.

Consider two related limits:

$$\lim_{(x,y)\to(0,0)} \cos(y) \qquad \text{and} \qquad \lim_{(x,y)\to(0,0)} \frac{\sin(x)}{x}.$$

The first limit does not contain x, and since $\cos(y)$ is continuous,

$$\lim_{(x,y)\to(0,0)} \cos(y)=\lim_{y\to 0}\cos(y)=\cos(0)=1.$$

The second limit does not contain y. By Theorem 8.6 we can say

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

Finally, following the properties of limits we can combine these two limits as follows:

$$\begin{aligned}\lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x} &= \lim_{(x,y)\to(0,0)}\left(\cos(y)\left(\frac{\sin(x)}{x}\right)\right)\\ &=\left(\lim_{(x,y)\to(0,0)}\cos(y)\right)\left(\lim_{(x,y)\to(0,0)}\frac{\sin(x)}{x}\right)\\ &=(1)(1)=1.\end{aligned}$$

We have found that

$$\lim_{(x,y)\to(0,0)} \frac{\cos(y)\sin(x)}{x}=f(0,0),$$

so f is continuous at $(0,0)$.

A similar analysis shows that f is continuous at all points in $\mathbb{R}^2$. As long as $x\neq 0$, we can evaluate the limit directly; when $x=0$, a similar analysis shows that the limit is $\cos(y)$. Thus we can say that f is continuous everywhere. A graph of f is given in Figure 16.7. Notice how it has no breaks, jumps, etc.

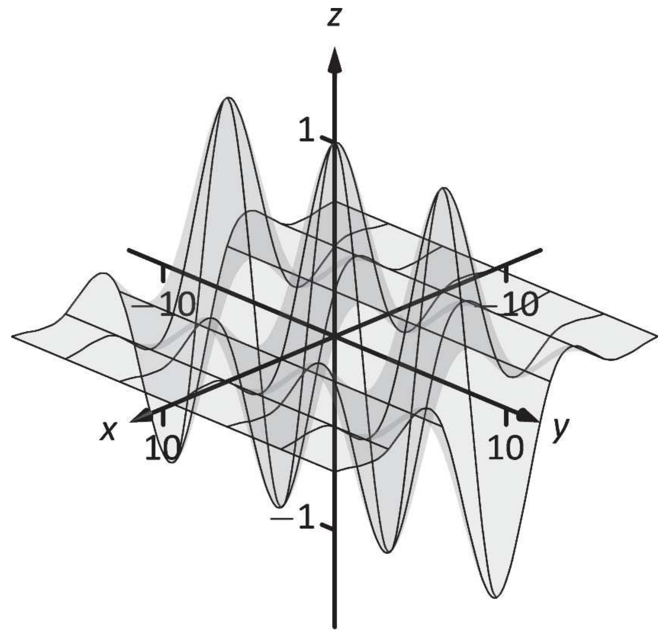


Figure 16.7: A graph of $f(x,y)$ in Example 16.7.

Of course, as with functions of one variable, we may combine continuous functions to create other continuous functions. More specifically, let f and g be continuous on a set S, let c be a real number, and let n be a positive integer. Then, the following functions are continuous on S.

- **Sums/Differences:** $f \pm g$
- **Constant multiples:** $c \cdot f$
- **Products:** $f \cdot g$
- **Quotients:** f/g (if $g \neq 0$ on S)
- **Powers:** f^n

For roots of a continuous function, we have that $\sqrt[n]{f}$ is continuous provided that $f \geq 0$ on S if n is even, whereas, if n is odd, this is true for all values of f on S. For what concerns function compositions, we let f be continuous on S, where the range of f on S is J, and let g be a single variable function that is continuous on J. Then $g \circ f$, i.e., $g(f(x,y))$, is continuous on S.

Having introduced the notion of continuity for functions of n variables, the multivariable counterpart of the indermediate value theorem follows intuitively.

Theorem 16.1 (Intermediate value theorem for functions of *n* variables)
Let f be a continuous function on D and, without loss of generality, let $f(\mathbf{a}) < f(\mathbf{b})$. Then for every value u, where $f(\mathbf{a}) < u < f(\mathbf{b})$, there is at least one interior point $\mathbf{c}$ in D such that $f(\mathbf{c}) = u$.

When considering single variable functions, we studied limits, then continuity, then the derivative. In our current study of multivariable functions, we have studied limits and continuity. In the next section we study derivation, which takes on a slight twist as we are in a multivarible context.

16.3 Partial derivatives

16.3.1 First partial derivatives

Let y be a function of x. We have studied in great detail the derivative of y with respect to x, that is, which measures the rate at which y changes with respect to x. Consider now $z = f(x, y)$. It makes sense to want to know how z changes with respect to x and/or y. This section begins our investigation into these rates of change.

Consider the function $z = f(x, y) = x^2 + 2y^2$, as graphed in Figure 16.8(a). By fixing $y = 2$, we focus our attention to all points on the surface where the y-value is 2, shown in Figures 16.8(a) and 16.8(b). These points form a curve in space: $z = f(x, 2) = x^2 + 8$ which is a function of just one variable. We can take the derivative of z with respect to x along this curve and find equations of tangent lines, etc.

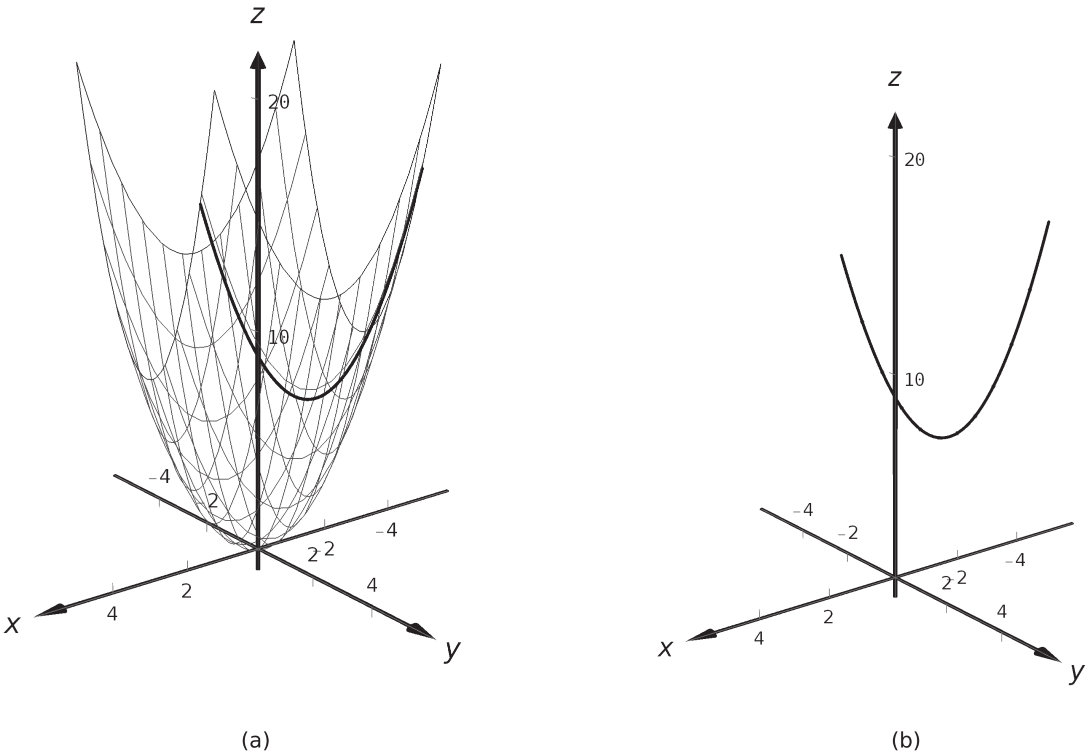


Figure 16.8: By fixing $y = 2$, the surface $f(x, y) = x^2 + 2y^2$ is a curve in space.

The key notion to extract from this example is: by treating y as constant (it does not vary) we can consider how z changes with respect to x. In a similar fashion, we can hold x constant and consider how z changes with respect to y. This is the underlying principle of **partial derivatives** (*partiële afgeleide*). We state the formal, limit–based definition first, then show how to compute these partial derivatives without directly taking limits.

Definition 16.7 (Partial derivative)
Let $z = f(x, y)$ be a continuous function on a set S in $\mathbb{R}^2$.

1. The **partial derivative of** f **with respect to** x is:

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

2. The **partial derivative of f with respect to** y is:

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

Alternate notations for $f_x(x,y)$ include:

$$\frac{\partial}{\partial x}f(x,y),\quad \frac{\partial f}{\partial x},\quad \frac{\partial z}{\partial x},\quad \text{and } z_x,$$

with similar notations for $f_y(x,y)$. For ease of notation, $f_x(x,y)$ is often abbreviated as f_x.

Example 16.8

Let $f(x,y) = x^2y+2x+y^3$. Find $f_x(x,y)$ using the limit definition.

Solution

Using Definition 16.7, we have:

$$\begin{aligned} f_x(x,y) &= \lim_{h\to 0} \frac{f(x+h,y)-f(x,y)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^2y+2(x+h)+y^3-(x^2y+2x+y^3)}{h} \\ &= \lim_{h\to 0} \frac{x^2y+2xhy+h^2y+2x+2h+y^3-(x^2y+2x+y^3)}{h} \\ &= \lim_{h\to 0} \frac{2xhy+h^2y+2h}{h} \\ &= \lim_{h\to 0} (2xy+hy+2) \\ &= 2xy+2. \end{aligned}$$

We have found $f_x(x,y) = 2xy+2$.

Using limits to compute partial derivatives is not necessary, as we can rely on our previous knowledge of derivatives to compute partial derivatives easily. When computing $f_x(x,y)$, we hold y fixed – it does not vary. Therefore we can compute the derivative with respect to x by treating y as a constant or coefficient.

Example 16.9

Find $f_x(x,y)$ and $f_y(x,y)$ for each of the following functions.

1. $f(x,y) = \cos(xy^2)+\sin(x)$
2. $f(x,y) = e^{x^2y^3}\sqrt{x^2+1}$

Solution

1. Begin with $f_x(x,y)$. We need to apply the chain rule with the cosine term; y^2 is the coefficient of the x-term inside the cosine function.

$$f_x(x,y) = -\sin(xy^2)(y^2)+\cos(x) = -y^2\sin(xy^2)+\cos(x).$$

To find $f_y(x,y)$, note that x is the coefficient of the y^2-term inside of the cosine term; also note that since x is fixed, $\sin(x)$ is also fixed, and we treat it as a constant.

$$f_y(x,y) = -\sin(xy^2)(2xy) = -2xy\sin(xy^2).$$

We may check our answer for what concerns f_x in Mathematica as follows:

```
In[27]:= D[Cos[x*y^2] + Sin[x], x]

Out[27]= Cos[x]-y^2 Sin[x y^2]
```

And likewise for what concerns f_y:

```
In[28]:= D[Cos[x*y^2] + Sin[x], y]

Out[28]= -2 x y Sin[x y^2]
```

2. Beginning with $f_x(x,y)$, note how we need to apply the product rule.

$$\begin{aligned} f_x(x,y) &= e^{x^2y^3}(2xy^3)\sqrt{x^2+1} + e^{x^2y^3}\frac{1}{2}(x^2+1)^{-1/2}(2x) \\ &= 2xy^3e^{x^2y^3}\sqrt{x^2+1} + \frac{xe^{x^2y^3}}{\sqrt{x^2+1}}. \end{aligned}$$

Note that when finding $f_y(x,y)$ we do not have to apply the product rule; since $\sqrt{x^2+1}$ does not contain y, we treat it as fixed and hence becomes a coefficient of the $e^{x^2y^3}$-term.

$$f_y(x,y) = e^{x^2y^3}(3x^2y^2)\sqrt{x^2+1} = 3x^2y^2e^{x^2y^3}\sqrt{x^2+1}.$$

We have shown how to compute a partial derivative, but it may still not be clear what a partial derivative means. Given $z = f(x,y)$, $f_x(x,y)$ measures the rate at which z changes as only x varies: y is held constant.

Imagine standing in a rolling meadow, then beginning to walk due east. Depending on your location, you might walk up, sharply down, or perhaps not change elevation at all. This is similar to measuring z_x: you are moving only east (in the x-direction) and not north/south at all. Going back to your original location, imagine now walking due north (in the y-direction). Perhaps walking due north does not change your elevation at all. This is analogous to $z_y = 0$: z does not change with respect to y. We can see that z_x and z_y do not have to be the same, or even similar, as it is easy to imagine circumstances where walking east means you walk downhill, though walking north makes you walk uphill.

16.3.2 Second partial derivatives

Let $z = f(x,y)$. We have learned to find the partial derivatives $f_x(x,y)$ and $f_y(x,y)$, which are each functions of x and y. Therefore we can take partial derivatives of them, each with respect to x and y. We define these second partials along with the notation, give examples, then discuss their meaning.

Definition 16.8 (Second and mixed partial derivative)
Let $z = f(x, y)$ be continuous on a set S.

1. The **second partial derivative of f with respect to x then x** (*tweede partiële afgeleide van f naar x*) is
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = \left(f_x\right)_x = f_{xx}.$$
2. The **second partial derivative of f with respect to x then y is** (*tweede partiële afgeleide van f naar x en y*) is
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = \left(f_x\right)_y = f_{xy}.$$

Similar definitions hold for $\dfrac{\partial^2 f}{\partial y^2} = f_{yy}$ and $\dfrac{\partial^2 f}{\partial x \partial y} = f_{yx}$. The second partial derivatives f_{xy} and f_{yx} are **mixed partial derivatives** (*gemengde partiële afgeleide*).

The terms in Definition 16.8 all depend on limits, so each definition comes with the caveat where the limit exists.

Example 16.10

For each of the following functions, find all 6 first and second partial derivatives. That is, find
$$f_x, \quad f_y, \quad f_{xx}, \quad f_{yy}, \quad f_{xy} \quad \text{and} \quad f_{yx}.$$

1. $f(x, y) = \dfrac{x^3}{y^2}$
2. $f(x, y) = e^x \sin(x^2 y)$

Solution

In each, we give f_x and f_y immediately and then spend time deriving the second partial derivatives.

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

$$f_x(x, y) = \frac{3x^2}{y^2}$$
$$f_y(x, y) = -\frac{2x^3}{y^3}$$
$$f_{xx}(x, y) = \frac{\partial}{\partial x}(f_x) = \frac{\partial}{\partial x}\left(\frac{3x^2}{y^2}\right) = \frac{6x}{y^2}$$
$$f_{yy}(x, y) = \frac{\partial}{\partial y}(f_y) = \frac{\partial}{\partial y}\left(-\frac{2x^3}{y^3}\right) = \frac{6x^3}{y^4}$$
$$f_{xy}(x, y) = \frac{\partial}{\partial y}(f_x) = \frac{\partial}{\partial y}\left(\frac{3x^2}{y^2}\right) = -\frac{6x^2}{y^3}$$
$$f_{yx}(x, y) = \frac{\partial}{\partial x}(f_y) = \frac{\partial}{\partial x}\left(-\frac{2x^3}{y^3}\right) = -\frac{6x^2}{y^3}$$

2. $f(x, y) = e^x \sin(x^2 y)$

   Because the following partial derivatives get rather long, we omit the extra notation and just give the results. In several cases, multiple applications of the product and chain rules will be necessary, followed by some basic combination of like terms.

   $f_x(x, y) = e^x \sin(x^2 y) + 2xye^x \cos(x^2 y)$

$$f_y(x,y) = x^2e^x\cos(x^2y)$$
$$f_{xx}(x,y) = e^x\sin(x^2y) + 4xye^x\cos(x^2y) + 2ye^x\cos(x^2y) - 4x^2y^2e^x\sin(x^2y)$$
$$f_{yy}(x,y) = -x^4e^x\sin(x^2y)$$
$$f_{xy}(x,y) = x^2e^x\cos(x^2y) + 2xe^x\cos(x^2y) - 2x^3ye^x\sin(x^2y)$$
$$f_{yx}(x,y) = x^2e^x\cos(x^2y) + 2xe^x\cos(x^2y) - 2x^3ye^x\sin(x^2y)$$

Higher-order partial derivatives can also be computed in Mathematica. For instance, given $f(x,y) = x^3/y^2$ from Example 16.10, we can find f_{xy} as follows:

```
In[29]:= D[x^3/y^2, x, y]

              2
           6 x
Out[29]= - ----
             3
            y
```

Notice how for each of the two functions in Example 16.10, $f_{xy} = f_{yx}$. Due to the complexity of the examples, this likely is not a coincidence. The following theorem states that it is not. It is also known as **Schwarz's theorem**, **Clairaut's theorem**, or **Young's theorem**.

Theorem 16.2 (Symmetry of second derivatives)
Let f be defined such that f_{xy} and f_{yx} are continuous on a set S. Then for each point (x,y) in S, $f_{xy}(x,y) = f_{yx}(x,y)$.

Now that we know how to find second partials, we investigate what they tell us.

Again we refer back to a function $y = f(x)$ of a single variable. The second derivative of f is "the derivative of the derivative," or "the rate of change of the rate of change." The second derivative measures how much the derivative is changing. If $f''(x) < 0$, then the derivative is getting smaller (so the graph of f is concave down); if $f''(x) > 0$, then the derivative is growing, making the graph of f concave up.

Now consider $z = f(x,y)$. Similar statements can be made about f_{xx} and f_{yy} as could be made about $f''(x)$ above. When taking derivatives with respect to x twice, we measure how much f_x changes with respect to x. If $f_{xx}(x,y) < 0$, it means that as x increases, f_x decreases, and the graph of f will be concave down in the x-direction. Using the analogy of standing in the rolling meadow used earlier in this section, f_{xx} measures whether one's path is concave up/down when walking due east. Similarly, f_{yy} measures the concavity in the y-direction. If $f_{yy}(x,y) > 0$, then f_y is increasing with respect to y and the graph of f will be concave up in the y-direction. Appealing to the rolling meadow analogy again, f_{yy} measures whether one's path is concave up/down when walking due north.

We now consider the mixed partials f_{xy} and f_{yx}. The mixed partial f_{xy} measures how much f_x changes with respect to y. Once again using the rolling meadow analogy, f_x measures the slope if one walks due east. Looking east, begin walking north (side–stepping). Is the path towards the east getting steeper? If so, $f_{xy} > 0$. Is the path towards the east not changing in steepness? If so, then $f_{xy} = 0$. A similar thing can be said about f_{yx}: consider the steepness of paths heading north while side–stepping to the east.

The following example examines these ideas with concrete numbers and graphs.

Example 16.11

Let $z = x^2 - y^2 + xy$. Evaluate the 6 first and second partial derivatives at $(-1/2, 1/2)$ and interpret what each of these numbers mean.

Solution

We find that:

$f_x(x,y) = 2x + y$, $f_y(x,y) = -2y + x$, $f_{xx}(x,y) = 2$, $f_{yy}(x,y) = -2$ and $f_{xy}(x,y) = f_{yx}(x,y) = 1$. Thus at $(-1/2, 1/2)$ we have

$$f_x\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}, \qquad f_y\left(-\frac{1}{2}, \frac{1}{2}\right) = -\frac{3}{2}.$$

The slope of the tangent line at $(-1/2, 1/2, -1/4)$ in the direction of x is $-1/2$: if one moves from that point parallel to the x-axis, the instantaneous rate of change will be $-1/2$. The slope of the tangent line at this point in the direction of y is $-3/2$: if one moves from this point parallel to the y-axis, the instantaneous rate of change will be $-3/2$. These tangents lines are graphed in Figure 16.9(a) and 16.9(b), together with the curve where $x = -1/2$ and $y = 1/2$, respectively, where the tangent lines are drawn in a solid line.

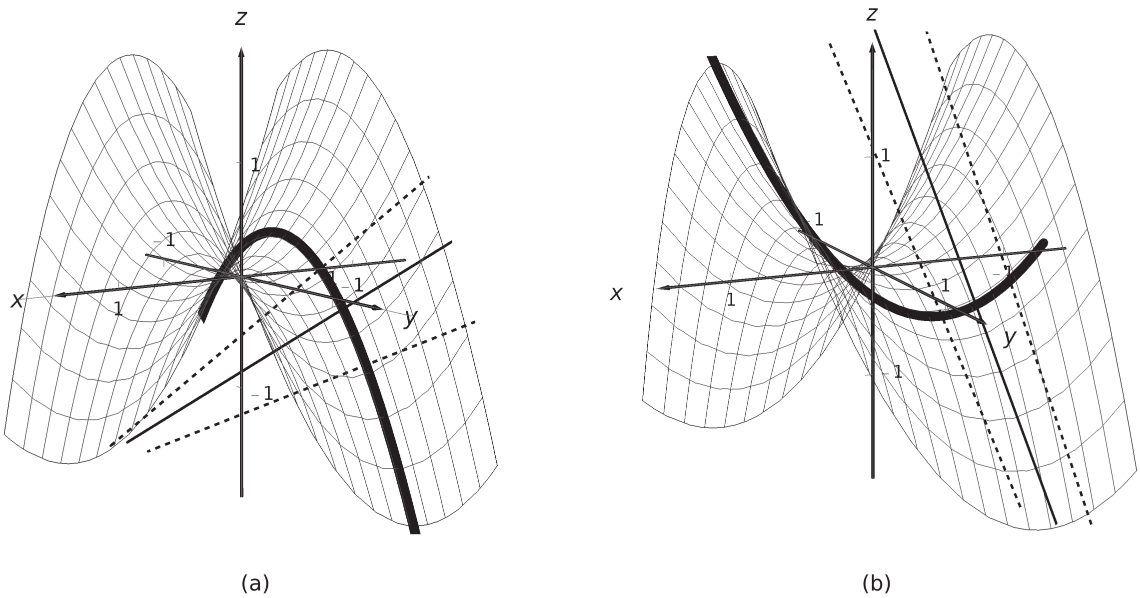


Figure 16.9: Understanding the second partial derivatives in Example 16.11.

Now consider only Figure 16.9(a). Three directed tangent lines are drawn (two are dashed), each in the direction of x; that is, each has a slope determined by f_x. Note how as y increases, the slope of these lines get closer to 0. Since the slopes are all negative, getting closer to 0 means the slopes are increasing. The slopes given by f_x are increasing as y increases, meaning f_{xy} must be positive.

Since $f_{xy} = f_{yx}$, we also expect f_y to increase as x increases. Consider Figure 16.9(b) where again three directed tangent lines are drawn, this time each in the direction of y with slopes determined by f_y. As x increases, the slopes become less steep (closer to 0). Since these are negative slopes, this means the slopes are increasing.

Thus far we have a visual understanding of f_x, f_y, and $f_{xy} = f_{yx}$. We now interpret f_{xx} and f_{yy}. In

Figure 16.9(a), we see a curve drawn where x is held constant at $x = -1/2$: only y varies. This curve is clearly concave down, corresponding to the fact that $f_{yy} < 0$. In Figure 16.9(b), we see a similar curve where y is constant and only x varies. This curve is concave up, corresponding to the fact that $f_{xx} > 0$.

16.3.3 Higher-order partial derivatives

Essentially, we can continue taking partial derivatives of partial derivatives of partial derivatives of . . . ; we do not have to stop with second partial derivatives. These higher order partial derivatives do not have a tidy graphical interpretation; nevertheless they are not hard to compute and worthy of some practice.

We do not formally define each higher order derivative, but rather give just a few examples of the notation. For instance,

$$f_{xyx}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right) \quad \text{and} \quad f_{xxz}(x,y) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\right).$$

Example 16.12

Let

$$f(x,y) = x^2y^2 + \sin(xy).$$

Find f_{xxy} and f_{yxx}.

Solution

To find f_{xxy}, we first find f_x, then f_{xx}, then f_{xxy}:

$$f_x = 2xy^2 + y\cos(xy)$$
$$f_{xx} = 2y^2 - y^2\sin(xy)$$

$$f_{xxy} = 4y - 2y\sin(xy) - xy^2\cos(xy).$$

To find f_{yxx}, we first find f_y, then f_{yx}, then f_{yxx}:

$$\begin{aligned} f_y &= 2x^2y + x\cos(xy) \\ f_{yx} &= 4xy + \cos(xy) - xy\sin(xy) \\ f_{yxx} &= 4y - y\sin(xy) - \big(y\sin(xy) + xy^2\cos(xy)\big) \\ &= 4y - 2y\sin(xy) - xy^2\cos(xy). \end{aligned}$$

Note how $f_{xxy} = f_{yxx}$.

In the previous example we saw that $f_{xxy} = f_{yxx}$; this is not a coincidence. While we do not state this as a formal theorem, as long as each partial derivative is continuous, it does not matter the order in which the partial derivatives are taken. For instance, $f_{xxy} = f_{xyx} = f_{yxx}$.

With $z = f(x,y)$, the partial derivatives f_x and f_y measure the instantaneous rate of change of z when moving parallel to the x- and y-axes, respectively. How do we measure the rate of change at a point when we do not move parallel to one of these axes? What if we move in the direction given by the vector $\langle 2,1\rangle$? Can we measure that rate of change? The answer is, of course, yes, we can. This is the topic of Section 16.6. First, we need to define what it means for a function of two variables to be differentiable.

16.3.4 Functions of n variables

The concepts underlying partial derivatives can be easily extend to n variables.

Definition 16.9 (Partial derivative with n variables)
Let $w = f(\mathbf{x})$ be a continuous function on a set D in $\mathbb{R}^n$.

The **partial derivative of f with respect to** x_i is:

$$f_{x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_i + h, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n)}{h}.$$

By taking partial derivatives of partial derivatives, we can find second partial derivatives of f with respect to z then y, for instance, just as before.

Example 16.13

For each of the following functions, find f_x, f_y, f_z, f_{xz}, f_{yz}, and f_{zz}.

1. $f(x,y,z) = x^2y^3z^4 + x^2y^2 + x^3z^3 + y^4z^4$
2. $f(x,y,z) = x\sin(yz)$

Solution

1. $f_x = 2xy^3z^4 + 2xy^2 + 3x^2z^3$
   $f_y = 3x^2y^2z^4 + 2x^2y + 4y^3z^4$
   $f_z = 4x^2y^3z^3 + 3x^3z^2 + 4y^4z^3$
   $f_{xz} = 8xy^3z^3 + 9x^2z^2$
   $f_{yz} = 12x^2y^2z^3 + 16y^3z^3$
   $f_{zz} = 12x^2y^3z^2 + 6x^3z + 12y^4z^2$

2. $f_x = \sin(yz)$
   $f_y = xz\cos(yz)$
   $f_z = xy\cos(yz)$
   $f_{xz} = y\cos(yz)$
   $f_{yz} = x\cos(yz) - xyz\sin(yz)$
   $f_{zz} = -xy^2\sin(xy)$

16.4 Total differential and differentiability

16.4.1 Total differential

We studied differentials in Section 9.7.3, where Definition 9.8 states that if $y = f(x)$ and f is differentiable, then $dy = f'(x)dx$. One important use of this differential is in integration by substitution. Another important application is approximation. Let $\Delta x = dx$ represent a change in x. When dx is small, $dy \approx \Delta y$, the change in y resulting from the change in x. So, as dx goes to 0, the error in approximating Δy with dy goes to 0.

We extend this idea to functions of two variables. Let $z = f(x,y)$, and let $\Delta x = dx$ and $\Delta y = dy$ represent changes in x and y, respectively (Figure 16.10). Let $\Delta z = f(x + dx, y + dy) - f(x,y)$ be the change in z over the change in x and y. Recalling that f_x and f_y give the instantaneous rates of z-change in the x- and y-directions, respectively, we can approximate Δz with $dz = f_x dx + f_y dy$; in words, the total change in z is approximately the change caused by changing x plus the change caused by changing y. In a moment we give an indication of whether or not this approximation is any good. First we give a name to dz.

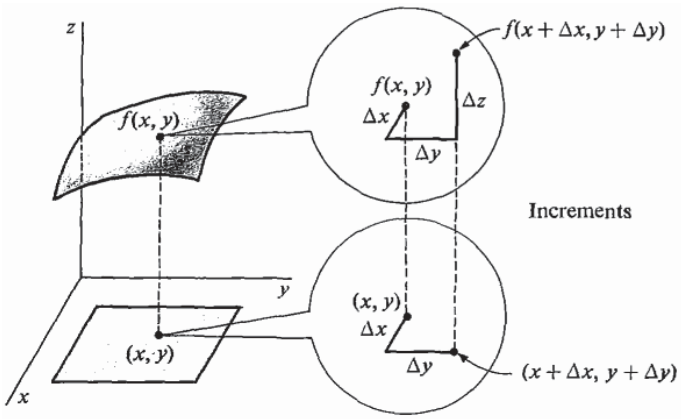


Figure 16.10: Understanding the total differential of a function of two variables.

Definition 16.10 (Total differential)
Let $z = f(x, y)$ be continuous on a set S. Let dx and dy represent changes in x and y, respectively. Where the partial derivatives f_x and f_y exist, the **total differential of** z (*totale differentiaal van* z) is

$$dz = f_x(x, y)dx + f_y(x, y)dy.$$

Note that from Definition 16.10, we can as well use vector notation:

$$dz = \left(f_x, f_y\right) \cdot (dx, dy).$$

16.4.2 Differentiability

16.4.2.1 Definition

We can approximate Δz with dz, but as with all approximations, there is error involved. A good approximation is one in which the error is small. At a given point (x_0, y_0), let E_x and E_y be functions of dx and dy such that $E_x dx + E_y dy$ describes this error. Then

$$\begin{aligned}\Delta z &= dz + E_x dx + E_y dy \\ &= f_x(x_0, y_0)dx + f_y(x_0, y_0)dy + E_x dx + E_y dy.\end{aligned}$$

If the approximation of Δz by dz is good, then as dx and dy get small, so does $E_x dx + E_y dy$. The approximation of Δz by dz is even better if, as dx and dy go to 0, so do E_x and E_y. This leads us to our definition of differentiability.

Definition 16.11 (Multivariable differentiability)
Let $z = f(x, y)$ be defined on a set S containing (x_0, y_0) where $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist. Let dz be the total differential of z at (x_0, y_0), let $\Delta z = f(x_0 + dx, y_0 + dy) - f(x_0, y_0)$, and let E_x and E_y be functions of dx and dy such that

$$\Delta z = dz + E_x dx + E_y dy.$$

1. We say f **is differentiable at** (x_0, y_0) (*afleidbaar*) if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||(dx, dy)|| < \delta$, then $||\left(E_x, E_y\right)|| < \varepsilon$. That is, as dx and dy go to 0, so do E_x and E_y.

2. We say f **is differentiable on** S (*afleidbaar over S*) if f is differentiable at every point in S. If f is differentiable on $\mathbb{R}^2$, we say that f is differentiable everywhere.

Example 16.14

Show $f(x,y) = xy + 3y^2$ is differentiable.

Solution

We begin by finding $f(x+dx, y+dy)$, Δz, f_x and f_y.

$$\begin{aligned} f(x+dx, y+dy) &= (x+dx)(y+dy) + 3(y+dy)^2 \\ &= xy + xdy + ydx + dxdy + 3y^2 + 6ydy + 3dy^2. \end{aligned}$$

$\Delta z = f(x+dx, y+dy) - f(x,y)$, so

$$\Delta z = xdy + ydx + dxdy + 6ydy + 3dy^2.$$

It is straightforward to compute $f_x = y$ and $f_y = x + 6y$. Consider once more Δz:

$$\begin{aligned} \Delta z &= xdy + ydx + dxdy + 6ydy + 3dy^2 \qquad \text{(Now reorder.)} \\ &= ydx + xdy + 6ydy + dxdy + 3dy^2 \\ &= \underbrace{(y)}_{f_x} dx + \underbrace{(x+6y)}_{f_y} dy + \underbrace{(dy)}_{E_x} dx + \underbrace{(3dy)}_{E_y} dy \\ &= f_x dx + f_y dy + E_x dx + E_y dy. \end{aligned}$$

With $E_x = dy$ and $E_y = 3dy$, it is clear that as dx and dy go to 0, E_x and E_y also go to 0. Since this did not depend on a specific point (x_0, y_0), we can say that $f(x,y)$ is differentiable for all pairs (x,y) in $\mathbb{R}^2$, or, equivalently, that f is differentiable everywhere.

Our intuitive understanding of differentiability of functions $y = f(x)$ of one variable was that the graph of f was **smooth** (*glad*). A similar intuitive understanding of functions $z = f(x,y)$ of two variables is that the surface defined by f is also smooth, not containing cusps, edges, breaks, etc. The following theorem states that differentiable functions are continuous, followed by another theorem that provides a more tangible way of determining whether a great number of functions are differentiable or not.

Theorem 16.3 (Continuity and differentiability of multivariable functions)
Let $z = f(x,y)$ be defined on a set S containing (x_0, y_0). If f is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

Theorem 16.4 (Differentiability of multivariable functions)
Let $z = f(x,y)$ be defined on a set S. If f_x and f_y are both continuous on S, then f is differentiable on S.

These theorems assure us that essentially all functions that we see in the course of our studies here are differentiable (and hence continuous) on their natural domains. There is a difference between Definition 16.11 and Theorem 16.4, though: it is possible for a function f to be differentiable yet f_x and/or f_y is not continuous. So when f_x and f_y exist at a point but are not continuous at that point, we need to use other methods to determine whether or not f is differentiable at that point.

For instance, consider the function

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

We can find $f_x(0,0)$ and $f_y(0,0)$ using Definition 16.7:

$$\begin{aligned} f_x(0,0) &= \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0; \\ f_y(0,0) &= \lim_{h\to 0} \frac{f(0,0+h)-f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{0}{h^2} = 0. \end{aligned}$$

Both f_x and f_y exist at $(0,0)$, but they are not continuous at $(0,0)$, as

$$f_x(x,y) = \frac{y(y^2-x^2)}{(x^2+y^2)^2} \qquad \text{and} \qquad f_y(x,y) = \frac{x(x^2-y^2)}{(x^2+y^2)^2}$$

are not continuous at $(0,0)$. Take the limit of f_x as $(x,y) \to (0,0)$ along the x- and y-axes; they give different results. So even though f_x and f_y exist at every point in the xy-plane, they are not continuous. Therefore it is possible, by Theorem 16.4, for f to not be differentiable.

Indeed, it is not. One can show that f is not continuous at $(0,0)$ (see Example 16.5), and by Theorem 16.3, this means f is not differentiable at $(0,0)$.

16.4.2.2 Approximating with differentials

By the definition, when f is differentiable dz is a good approximation for Δz when dx and dy are small. We give a simple example of how this is used here. We can use this to approximate error propagation; that is, if the input is a little off from what it should be, how far from correct will the output be? We demonstrate this in an example.

Example 16.15

A cylindrical steel storage tank is to be built that is 10m tall and 4m across in diameter. It is known that the steel will expand/contract with temperature changes; is the overall volume of the tank more sensitive to changes in the diameter or in the height of the tank?

Solution

A cylindrical solid with height h and radius r has volume $V = \pi r^2 h$. We can view V as a function of two variables, r and h. We can compute partial derivatives of V:

$$\frac{\partial V}{\partial r} = V_r(r,h) = 2\pi rh \qquad \text{and} \qquad \frac{\partial V}{\partial h} = V_h(r,h) = \pi r^2.$$

The total differential is $dV = (2\pi rh)dr + (\pi r^2)dh$. When $h = 10$ and $r = 2$, we have
$dV = 40\pi dr + 4\pi dh$. Note that the coefficient of dr is $40\pi \approx 125.7$; the coefficient of dh is a tenth of that, approximately 12.57. A small change in radius will be multiplied by 125.7, whereas a

small change in height will be multiplied by 12.57. Thus the volume of the tank is more sensitive to changes in radius than in height.

The previous example showed that the volume of a particular tank was more sensitive to changes in radius than in height. Keep in mind that this analysis only applies to a tank of those dimensions. A tank with a height of 0.3 m and radius of 5 m would be more sensitive to changes in height than in radius. One could make a chart of small changes in radius and height and find exact changes in volume given specific changes.

16.4.3 Functions of n variables

The definition of differentiability for functions of n variables is very similar to that of functions of two variables. We again start with the total differential.

Definition 16.12 (Total differential)
Let $w = f(\mathbf{x})$ be continuous on a set D. Let dx_i represent change in x_i. Where the partial derivatives f_i, $i = 1, \ldots, n$, exist, **the total differential of** w is

$$dw = \sum_{i=1}^{n} f_{x_i}(\mathbf{x}) dx_i.$$

Of course, assuming that we stick to the same increments dx_i, it is relatively straightforward to see that we can extend this definition to higher-order differentials.

Definition 16.13 (The n-th order total differential)
Let $w = f(\mathbf{x})$ be continuous on a set D. Let dx_i represent change in x_i. Where the partial derivatives f_i, $i = 1, \ldots, n$, exist, **the n-th order total differential of** w is

$$d^n w = \left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x}) dx_i \right)^n .$$

To understand this definition correctly, it is important to realize that $\frac{\partial^n}{\partial x_i^n}$ represents the n-th power of $\frac{\partial}{\partial x_i}$ when expanding the power n according to the binomial theorem.

The first-order total differential given in Definition 16.12 can be a good approximation of the change in w when $w = f(\mathbf{x})$ is differentiable.

Definition 16.14 (Multivariable differentiability)
Let $w = f(\mathbf{x})$ be defined on a set D containing $\mathbf{c}$ where $f_{x_i}(\mathbf{c})$, $i = 1, \ldots, n$ exist. Let dw be the total differential of w at $\mathbf{c}$, let $\Delta w = f(c_1 + dx_1, c_2 + dx_2, \ldots, c_n + dx_n) - f(c_1, c_2, \ldots, c_n)$
$= f(\mathbf{c} + \mathbf{dx}) - f(\mathbf{c})$, and let E_{x_i} be functions of dx_i for $i = 1, \ldots, n$ such that

$$\Delta w = dw + \sum_{i=1}^{n} E_{x_i} dx_i.$$

1. We say f **is differentiable at c** if, given $\varepsilon > 0$, there is a $\delta > 0$ such that if $||\mathbf{dx}|| < \delta$, then $||\mathbf{E_x}|| < \varepsilon$.

2. We say f is differentiable on B if f is differentiable at every point in B. If f is differentiable on $\mathbb{R}^n$, we say that f is differentiable everywhere.

Just as before, this definition gives a rigorous statement about what it means to be differentiable that is not very intuitive. We follow it with a theorem similar to Theorem 16.4.

Theorem 16.5 (Continuity and differentiability of functions of three variables)
Let $w = f(\mathbf{x})$ be defined on a set D containing $\mathbf{c}$.

1. *If f is differentiable at $\mathbf{c}$, then f is continuous at $\mathbf{c}$.*
2. *If the f_{x_i}, $i = 1,\ldots,n$ are continuous on B, then f is differentiable on B.*

16.5 The multivariable chain rule and implicit function theorem

16.5.1 Rationale

Consider driving an off-road vehicle along a dirt road. As you drive, your elevation likely changes. What factors determine how quickly your elevation rises and falls? After some thought, generally one recognizes that one's velocity (speed and direction) and the terrain influence your rise and fall.

One can represent the terrain as the surface defined by a multivariable function $z = f(x,y)$; one can represent the path of the off-road vehicle, as seen from above, with a vector–valued function $\vec{r}(t) = \big(x(t),y(t)\big)$; the velocity of the vehicle is thus $\vec{r}\,'(t) = \big(x'(t),y'(t)\big)$.

Consider Figure 16.11 in which a surface $z = f(x,y)$ is drawn, along with a dashed curve in the xy-plane. Restricting f to just the points on this circle gives the curve shown on the surface (i.e., the path of the vehicle.) The derivative $\frac{df}{dt}$ gives the instantaneous rate of change of f with respect to t. If we consider an object travelling along this path, $\frac{df}{dt} = \frac{dz}{dt}$ gives the rate at which the object rises/falls. Conceptually, the multivariable chain rule combines terrain and velocity information properly to compute this rate of elevation change.

Abstractly, let z be a function of x and y; that is, $z = f(x,y)$ for some function f, and let x and y each be functions of t. By choosing a t-value, x- and y-values are determined, which in turn determine z: this defines z as a function of t. The multivariable chain rule gives a method of computing $\frac{dz}{dt}$.

Theorem 16.6 (Multivariable chain rule, Part I)
Let $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, where f, g and h are differentiable functions. Then $z = f(x,y) = f\big(g(t),h(t)\big)$ is a function of t, and

$$\begin{aligned}\frac{dz}{dt} = \frac{df}{dt} &= f_x(x,y)\frac{dx}{dt} + f_y(x,y)\frac{dy}{dt}\\ &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \qquad (16.1)\\ &= \big(f_x,f_y\big)\cdot\big(x',y'\big).\end{aligned}$$

Although this theorem should be clear if one has a good understanding of the chain rule, which forces you to work from outer function to the inner one, it can be shown more formally, for instance, by

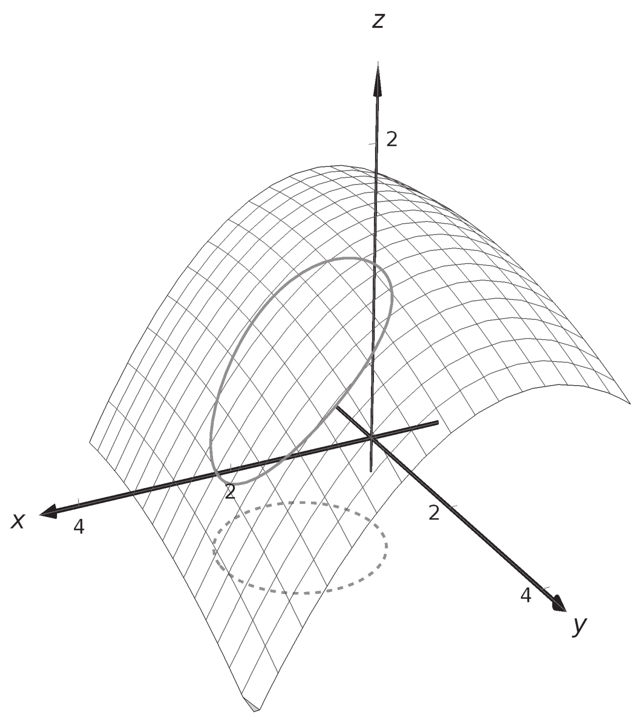


Figure 16.11: Understanding the application of the multivariable chain rule.

resorting to Definition 16.10, which allows us to write that

$$dz = f_x(x,y)\,dx + f_y(x,y)\,dy\,.$$

Dividing both sides by dt and recalling that $x = g(t)$ and $y = h(t)$ are differentiable functions, we get

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}\,.$$

Notice, the third line of equations in Theorem 16.6. The vector (f_x, f_y) contains information about the surface (terrain); the vector (x', y') can represent velocity. In the context measuring the rate of elevation change of the off-road vehicle, the multivariable chain rule states it can be found through a product of terrain and velocity information.

We now practice applying the multivariable chain rule.

Example 16.16

Let $z = f(x,y) = x^2y + x$, where $x = \sin(t)$ and $y = e^{5t}$. Find $\frac{dz}{dt}$ using the chain rule.

Solution

Following Theorem 16.6, we first find

$$f_x(x,y) = 2xy + 1, \qquad f_y(x,y) = x^2, \qquad \frac{dx}{dt} = \cos(t), \qquad \frac{dy}{dt} = 5e^{5t}.$$

Applying the theorem, we have

$$\frac{dz}{dt} = (2xy+1)\cos(t) + 5x^2e^{5t}.$$

This may look odd, as it seems that $\frac{dz}{dt}$ is a function of x, y and t. Since x and y are functions of t, $\frac{dz}{dt}$ is really just a function of t, and we can replace x with $\sin(t)$ and y with e^{5t} to arrive of:

$$\frac{dz}{dt} = (2\sin(t)e^{5t}+1)\cos(t)+5e^{5t}\sin^2(t).$$

The previous example can make us wonder: if we substituted for x and y at the end to show that $\frac{dz}{dt}$ is really just a function of t, why not substitute before differentiating, showing clearly that z is a function of t?

That is, $z = x^2y+x = \sin^2(t)e^{5t}+\sin(t)$. Applying the chain and product rules, we have

$$\frac{dz}{dt} = 2\sin(t)\cos(t)\,e^{5t}+5\sin^2(t)\,e^{5t}+\cos(t),$$

which matches the result from the example.

This may now make one wonder "What's the point? If we could already find the derivative, why learn another way of finding it?" In some cases, applying this rule makes deriving simpler, but this is hardly the power of the chain rule. Rather, in the case where $z = f(x,y)$, $x = g(t)$ and $y = h(t)$, the chain rule is extremely powerful when we do not know what f, g and/or h are. We demonstrate this in the next example.

Example 16.17

An object travels along a path on a surface. The exact path and surface are not known, but at time $t = t_0$ it is known that :

$$\frac{\partial z}{\partial x} = 5, \qquad \frac{\partial z}{\partial y} = -2, \qquad \frac{dx}{dt} = 3 \qquad \text{and} \qquad \frac{dy}{dt} = 7.$$

Find $\frac{dz}{dt}$ at time t_0.

Solution

The multivariable chain rule states that

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt}+\frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= 5(3)+(-2)(7)\\ &= 1.\end{aligned}$$

By knowing certain rates–of–change information about the surface and about the path of the particle in the xy-plane, we can determine how quickly the object is rising/falling.

Of course, we might as well be interested in the second derivative of $z = f(x,y)$. To compute that derivative, it is important to realize that one again gets a function of t upon substituting $x = g(t)$ and $y = h(t)$ in the right-hand side of Equation (16.1). So we may compute

$$\frac{d^2z}{dt^2} = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\frac{dx}{dt}+\frac{\partial f}{\partial y}\frac{dy}{dt}\right).$$

Applying the product rule, we get

$$\frac{d^2z}{dt^2} = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\right)\frac{dx}{dt}+\frac{\partial f}{\partial x}\frac{d^2x}{dt^2}+\frac{d}{dt}\left(\frac{\partial f}{\partial y}\right)\frac{dy}{dt}+\frac{\partial f}{\partial y}\frac{d^2y}{dt^2},$$

and after expanding the derivatives in the first and third term:

$$\frac{d^2z}{dt^2} = \left(\frac{\partial^2 f}{\partial x^2}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y \partial x}\frac{dy}{dt}\right)\frac{dx}{dt} + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \left(\frac{\partial^2 f}{\partial x \partial y}\frac{dx}{dt} + \frac{\partial^2 f}{\partial y^2}\frac{dy}{dt}\right)\frac{dy}{dt} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2}.$$

After simplification, we arrive at

$$\frac{d^2z}{dt^2} = \frac{\partial^2 f}{\partial x^2}\left(\frac{dx}{dt}\right)^2 + 2\frac{\partial^2 f}{\partial x \partial y}\frac{dx}{dt}\frac{dy}{dt} + \frac{\partial^2 f}{\partial y^2}\left(\frac{dy}{dt}\right)^2 + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2},$$

which we can write as

$$\frac{d^2z}{dt^2} = \left(\frac{\partial}{\partial x}\frac{dx}{dt} + \frac{\partial}{\partial y}\frac{dy}{dt}\right)^2 f + \frac{\partial f}{\partial x}\frac{d^2x}{dt^2} + \frac{\partial f}{\partial y}\frac{d^2y}{dt^2}.$$

A similar reasoning is possible for higher-order derivatives. Note that, again, just as in Definition 16.13, $\left(\frac{\partial}{\partial x}\right)^2$ is to be understood as $\frac{\partial^2}{\partial x^2}$.

We can also extend the chain rule to include the situation where z is a function of more than one variable, and each of these variables is also a function of more than one variable. The basic case of this is where $z = f(x,y)$, and x and y are functions of two variables, say s and t.

Theorem 16.7 (Multivariable chain rule, Part II)

1. *Let $z = f(x,y)$, $x = g(s,t)$ and $y = h(s,t)$, where f, g and h are differentiable functions. Then z is a function of s and t, and*

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

2. *Let $z = f(\mathbf{x})$ be a differentiable function of n variables, where each of the x_i is a differentiable function of the variables $t_1, t_2, \ldots, t_n$. Then z is a function of the t_j, and*

$$\frac{\partial z}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j}.$$

Example 16.18

Let $z = f(x,y) = x^2y + x$, $x = s^2 + 3t$ and $y = 2s - t$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$, and evaluate each when $s = 1$ and $t = 2$.

Solution

Following Theorem 16.7, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = 2xy + 1 \qquad \frac{\partial f}{\partial y} = x^2,$$

$$\frac{\partial x}{\partial s} = 2s \qquad \frac{\partial x}{\partial t} = 3 \qquad \frac{\partial y}{\partial s} = 2 \qquad \frac{\partial y}{\partial t} = -1.$$

Thus

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} = (2xy+1)(2s) + (x^2)(2) = 4xys + 2s + 2x^2,$$

and

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = (2xy+1)(3) + (x^2)(-1) = 6xy - x^2 + 3.$$

When $s = 1$ and $t = 2$, $x = 7$ and $y = 0$, so

$$\frac{\partial z}{\partial s} = 100 \qquad \text{and} \qquad \frac{\partial z}{\partial t} = -46.$$

Example 16.19

Let $w = xy + z^2$, where $x = t^2e^s$, $y = t\cos(s)$, and $z = s\sin(t)$. Find $\frac{\partial w}{\partial t}$ when $s = 0$ and $t = \pi$.

Solution

Following Theorem 16.7, we compute the following partial derivatives:

$$\frac{\partial f}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \qquad \frac{\partial f}{\partial z} = 2z,$$

$$\frac{\partial x}{\partial t} = 2te^s \qquad \frac{\partial y}{\partial t} = \cos(s) \qquad \frac{\partial z}{\partial t} = s\cos(t).$$

Thus

$$\frac{\partial w}{\partial t} = y(2te^s) + x(\cos(s)) + 2z(s\cos(t)).$$

When $s = 0$ and $t = \pi$, we have $x = \pi^2$, $y = \pi$ and $z = 0$. Thus

$$\frac{\partial w}{\partial t} = \pi(2\pi) + \pi^2 = 3\pi^2.$$

As indicated before, the real strength of the multivariable chain rule lies in the fact that one can compute derivatives and differentials of multivariable functions even not knowing the underlying function definitions. This is illustrated in the following example.

Example 16.20

In each of the following cases, determine dz and d^2z

1. $z = f(u)$ and $u = g(x,y)$
2. $z = f(u,v)$, $u = g(x,y)$ and $v = h(x,y)$

Solution

1. (a)
$$\begin{aligned} dz &= \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \\ &= \frac{df}{du}\frac{\partial u}{\partial x}\,dx + \frac{df}{du}\frac{\partial u}{\partial y}\,dy \\ &= \frac{df}{du}\left[\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right] \\ &= \frac{df}{du}\,du \end{aligned}$$

$$\begin{aligned}
\text{(b) } d^2z &= \frac{\partial^2 z}{\partial x^2}\,dx^2 + 2\frac{\partial^2 z}{\partial x\partial y}\,dx\,dy + \frac{\partial^2 z}{\partial y^2}\,dy^2 \\
&= \left[\frac{d^2f}{du^2}\left(\frac{\partial u}{\partial x}\right)^2 + \frac{df}{du}\,\frac{\partial^2 u}{\partial x^2}\right]dx^2 + 2\left[\frac{d^2f}{du^2}\,\frac{\partial u}{\partial x}\,\frac{\partial u}{\partial y} + \frac{df}{du}\,\frac{\partial^2 u}{\partial x\partial y}\right]dx\,dy \\
&\qquad + \left[\frac{d^2f}{du^2}\left(\frac{\partial u}{\partial y}\right)^2 + \frac{df}{du}\,\frac{\partial^2 u}{\partial y^2}\right]dy^2 \\
&= \frac{d^2f}{du^2}\left[\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right]^2 + \frac{df}{du}\left[\frac{\partial}{\partial x}\,dx + \frac{\partial}{\partial y}\,dy\right]^2 u \\
&= \frac{d^2f}{du^2}\,du^2 + \frac{df}{du}\,d^2u
\end{aligned}$$

2.
$$\begin{aligned}
\text{(a) } dz &= \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \\
&= \left(\frac{\partial f}{\partial u}\,\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\,\frac{\partial v}{\partial x}\right)dx + \left(\frac{\partial f}{\partial u}\,\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\,\frac{\partial v}{\partial y}\right)dy \\
&= \frac{\partial f}{\partial u}\left(\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right) + \frac{\partial f}{\partial v}\left(\frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy\right) \\
&= \frac{\partial f}{\partial u}\,du + \frac{\partial f}{\partial v}\,dv
\end{aligned}$$

$$\begin{aligned}
\text{(b) } d^2z &= \frac{\partial^2 z}{\partial x^2}\,dx^2 + 2\,\frac{\partial^2 z}{\partial x\partial y}\,dx\,dy + \frac{\partial^2 z}{\partial y^2}\,dy^2 \\
&= \left[\frac{\partial^2 f}{\partial u^2}\left(\frac{\partial u}{\partial x}\right)^2 + 2\,\frac{\partial^2 f}{\partial u\partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial^2 f}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2 + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial x^2} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial x^2}\right]dx^2 \\
&\quad + 2\left[\frac{\partial^2 f}{\partial u^2}\,\frac{\partial u}{\partial x}\,\frac{\partial u}{\partial y} + \frac{\partial^2 f}{\partial u\partial v}\,\frac{\partial u}{\partial x}\,\frac{\partial v}{\partial y} + \frac{\partial^2 f}{\partial u\partial v}\,\frac{\partial u}{\partial y}\,\frac{\partial v}{\partial x}\right. \\
&\qquad\qquad \left. + \frac{\partial^2 f}{\partial v^2}\,\frac{\partial v}{\partial x}\,\frac{\partial v}{\partial y} + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial x\partial y} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial x\partial y}\right]dxdy \\
&\quad + \left[\frac{\partial^2 f}{\partial u^2}\left(\frac{\partial u}{\partial y}\right)^2 + 2\frac{\partial^2 f}{\partial u\partial v}\frac{\partial u}{\partial y}\frac{\partial v}{\partial y} + \frac{\partial^2 f}{\partial v^2}\left(\frac{\partial v}{\partial y}\right)^2 + \frac{\partial f}{\partial u}\frac{\partial^2 u}{\partial y^2} + \frac{\partial f}{\partial v}\frac{\partial^2 v}{\partial y^2}\right]dy^2 \\
&= \frac{\partial^2 f}{\partial u^2}\left[\left(\frac{\partial u}{\partial x}\right)^2 dx^2 + 2\,\frac{\partial u}{\partial x}\,\frac{\partial u}{\partial y}\,dx\,dy + \left(\frac{\partial u}{\partial y}\right)^2 dy^2\right] \\
&\quad + 2\,\frac{\partial^2 f}{\partial u\partial v}\left[\frac{\partial u}{\partial x}\,\frac{\partial v}{\partial x}\,dx^2 + \frac{\partial u}{\partial x}\,\frac{\partial v}{\partial y}\,dx\,dy + \frac{\partial u}{\partial y}\,\frac{\partial v}{\partial x}\,dx\,dy + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y}\,dy^2\right] \\
&\quad + \frac{\partial^2 f}{\partial v^2}\left[\left(\frac{\partial v}{\partial x}\right)^2 dx^2 + 2\,\frac{\partial v}{\partial x}\,\frac{\partial v}{\partial y}\,dx\,dy + \left(\frac{\partial v}{\partial y}\right)^2 dy^2\right] \\
&\quad + \frac{\partial f}{\partial u}\left[\frac{\partial^2 u}{\partial x^2}\,dx^2 + 2\,\frac{\partial^2 u}{\partial x\partial y}\,dx\,dy + \frac{\partial^2 u}{\partial y^2}\,dy^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{\partial f}{\partial v}\left[\frac{\partial^2 v}{\partial x^2}\, dx^2 + 2\,\frac{\partial^2 v}{\partial x \partial y}\, dx\, dy + \frac{\partial^2 v}{\partial y^2}\, dy^2\right] \\
= &\ \frac{\partial f^2}{\partial u^2}\, du^2 + 2\frac{\partial^2 f}{\partial u \partial v}\, du\, dv + \frac{\partial^2 f}{\partial v^2}\, dv^2 + \frac{\partial f}{\partial u}\, d^2 u + \frac{\partial f}{\partial v}\, d^2 v \\
= &\ \left[\frac{\partial}{\partial u}\, du + \frac{\partial}{\partial v}\, dv\right]^2 f + \frac{\partial f}{\partial u}\, d^2 u + \frac{\partial f}{\partial v}\, d^2 v
\end{aligned}
$$

16.5.2 The implicit function theorem

We studied finding $\frac{dy}{dx}$ when y is given as an implicit function of x in detail in Section 9.4. We find here that the multivariable chain rule gives a simpler method of finding $\frac{dy}{dx}$.

For instance, consider the implicit function $x^2y - xy^3 = 3$. We learned to use the following steps to find $\frac{dy}{dx}$:

$$
\begin{aligned}
& \frac{d}{dx}\left(x^2y - xy^3\right) = \frac{d}{dx}\left(3\right) \\
\Rightarrow \quad & 2xy + x^2\frac{dy}{dx} - y^3 - 3xy^2\frac{dy}{dx} = 0 \\
\Leftrightarrow \quad & \frac{dy}{dx} = -\frac{2xy - y^3}{x^2 - 3xy^2}.
\end{aligned}
\tag{16.2}
$$

Instead of using this method, consider $z = x^2y - xy^3$. The implicit function above describes the level curve $z = 3$. Considering x and y as functions of x, the multivariable chain rule states that

$$\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx}. \tag{16.3}$$

Since z is constant (in our example, $z = 3$) it holds that, $\frac{dz}{dx} = 0$. We also know $\frac{dx}{dx} = 1$. Consequently, equation (16.3) becomes

$$0 = \frac{\partial z}{\partial x}(1) + \frac{\partial z}{\partial y}\frac{dy}{dx}.$$

Consequently,

$$\frac{dy}{dx} = -\frac{\partial z}{\partial x}\bigg/\frac{\partial z}{\partial y} = -\frac{f_x}{f_y} \tag{16.4}$$

Note how our solution for $\frac{dy}{dx}$ in Equation (16.2) is just the partial derivative of z with respect to x, divided by the partial derivative of z with respect to y, all multiplied by (-1).

Actually, Equation (16.4) is a consequence of the powerful implicit function theorem (*impliciete functiestelling*), which can be formulated as follows for implicit functions of two variables $F(x, y) = 0$.

Theorem 16.8 (The implicit function theorem)
Let F be a continuously differentiable implicit function of two variables x and y, i.e. it is of differentiability class C^1, let (x_0, y_0) be a point for which $F(x_0, y_0) = 0$ and let

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0\,.$$

Then there exists an open interval I containing x_0 and a unique C^1 function $f : I \mapsto \mathbb{R}$ such that $f(x_0) = y_0$ and $F(x, f(x)) = 0$ for all $x \in I$.

Essentially, the implicit function theorem allows relations

$$\left\{(x,y) \in \mathbb{R}^2 \mid F(x,y) = 0\right\}$$

to be converted to functions of a real variable. It does so by representing the relation as the graph of a function. There may, however, not be a single function whose graph can represent the entire relation, but there may be such a function on a restriction of the domain of the relation. The implicit function theorem gives a sufficient condition to ensure that there is such a function.

Example 16.21

Let us consider $F(x,y) = x^2 + y^2 - 1 = 0$. We know that the equation $x^2 + y^2 - 1 = 0$ cuts out the unit circle. There is no way to represent the unit circle as the graph of a function of one variable $y = f(x)$ because for each choice of $x \in [-1, 1]$, there are two choices of y, namely $y = \pm\sqrt{1-x^2}$. It even works in situations where we do not have a formula for $F(x,y)$.

Still, as long as $y_0 \neq 0$, the conditions of Theorem 16.8 are fulfilled and we are guaranteed that there exists a unique function f of one variable x on $]-1,1[$ for which $f(x_0) = y_0$. More specifically, for $-1 < x_0 < 1$ and $y_0 > 0$, we let $f_1(x) = \sqrt{1-x^2}$ and the graph of $y = f_1(x)$ provides the upper half of the circle. Similarly, for $-1 < x_0 < 1$ and $y_0 < 0$, we let $f_2(x) = -\sqrt{1-x^2}$ and the graph of $y = f_2(x)$ gives the lower half of the circle. So, it is possible to represent part of the circle as the graph of a function of one variable.

At the intersections of the unit circle with x-axis, i.e. where $y = 0$, however, we observe that

$$\frac{\partial F}{\partial y} = 0\,,$$

which implies that the conditions in Theorem 16.8 are not met and we hence may not use it. We observe that these points of intersection lie on the graphs of both f_1 and f_2 when these are extended to the closed interval $[-1, 1]$.

The purpose of the implicit function theorem is to tell us the existence of functions like $f_1(x)$ and $f_2(x)$, even in situations where we cannot write them down with explicit formulas. The theorem guarantees that $f_1(x)$ and $f_2(x)$ are differentiable. For instance, for what concerns the implicit function at stake, we may take advantage of the fact that $F(x, f_i(x)) = 0$ as long as $y_0 \neq 0$ to write

$$\begin{aligned} \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{df_i}{dx} &= 0\,, \qquad (16.5)\\ \Leftrightarrow \quad 2x + 2y\frac{df_i}{dx} &= 0\,, \end{aligned}$$

from which we conclude that

$$\frac{df_i}{dx} = -\frac{x}{y}.$$

Consequently, we find that

$$\frac{df_1}{dx} = -\frac{x}{\sqrt{1-x^2}}$$

or

$$\frac{df_2}{dx} = \frac{x}{\sqrt{1-x^2}}.$$

Of course, the second-order derivatives can computed by applying the chain rule once more to Equation (16.5) to arrive at

$$\frac{\partial^2 F}{\partial x^2} + 2\frac{\partial^2 F}{\partial x \partial y}\frac{df_i}{dx} + \frac{\partial F}{\partial y}\frac{d^2 f_i}{dx^2} + \frac{\partial^2 F}{\partial y^2}\left(\frac{df_i}{dx}\right)^2 = 0.$$

From this expression, it is easy to compute $\frac{d^2 f_i}{dx^2}$.

Having introduced the implicit function theorem for functions of two variables, we now extend it to implicit functions of $n+1$ variables $F(\mathbf{x}, y)$.

Theorem 16.9 (The general implicit function theorem)
Let F be a continuously differentiable implicit function of $n+1$ variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and y, i.e. it is of differentiability class C^1, let $(\mathbf{x}_0, y_0)$ be a point for which $F(\mathbf{x}_0, y_0) = 0$ and let

$$\frac{\partial F}{\partial y}(\mathbf{x}_0, y_0) \neq 0.$$

Then there exists an open n—dimensional ball B containing $\mathbf{x}_0$ and a unique C^1 function $f : B \mapsto \mathbb{R}$ such that $f(\mathbf{x}_0) = y_0$ and $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $\mathbf{x} \in B$.

To settle the mind, let us consider one more example.

Example 16.22

Consider the following implicit function of three variables

$$F(x, y, z) = ze^z - x - 3y = 0.$$

Clearly, near the point $(e, 0, 1)$ the conditions of Theorem 16.9 are fulfilled since $F(e, 0, 1) = 0$ and

$$\frac{\partial F}{\partial z}(e, 0, 1) = 2e.$$

So, there exists a function $z = f(x, y)$, and we may proceed by computing its first-order derivatives as follows

$$\begin{aligned} \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial f}{\partial x} = -1 + (z+1)e^z \frac{\partial f}{\partial x} &= 0, \\ \frac{\partial F}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial f}{\partial y} = -3 + (z+1)e^z \frac{\partial f}{\partial y} &= 0, \end{aligned} \tag{16.6}$$

from which we obtain

$$\frac{\partial f}{\partial x} = \frac{e^{-z}}{z+1} \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{3e^{-z}}{z+1}.$$

Taking the derivatives of the expressions in Equation (16.6) yields

$$\begin{aligned}
(z+2)e^z\left(\frac{\partial f}{\partial x}\right)^2 + (z+1)e^z\frac{\partial^2 f}{\partial x^2} &= 0,\\
(z+2)e^z\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} + (z+1)e^z\frac{\partial^2 f}{\partial x \partial y} &= 0,\\
(z+2)e^z\left(\frac{\partial f}{\partial y}\right)^2 + (z+1)e^z\frac{\partial^2 f}{\partial y^2} &= 0,
\end{aligned}$$

such that

$$\frac{\partial^2 f}{\partial x^2} = -\frac{(z+2)e^z}{(z+1)e^z}\left(\frac{\partial f}{\partial x}\right)^2 = -\frac{(z+2)e^{-2z}}{(z+1)^3}.$$

In Section 16.3 we learned how partial derivatives give certain instantaneous rate of change information about a function $z = f(x,y)$. In that section, we measured the rate of change of f by holding one variable constant and letting the other vary. We can visualize this change by considering the surface defined by f at a point and moving parallel to the x-axis.

What if we want to move in a direction that is not parallel to a coordinate axis? Can we still measure instantaneous rates of change? Yes; we find out how in the next section. In doing so, we will see how the multivariable chain rule informs our understanding of these directional derivatives.

16.6 Directional derivatives

16.6.1 Definition

Partial derivatives give us an understanding of how a surface changes when we move in the x- and y-directions. We made the comparison to standing in a rolling meadow and heading due east: the amount of rise/fall in doing so is comparable to f_x. Likewise, the rise/fall in moving due north is comparable to f_y. The steeper the slope, the greater in magnitude f_y.

But what if we did not move due north or east? What if we needed to move northeast and wanted to measure the amount of rise/fall? Partial derivatives alone cannot measure this. This section investigates **directional derivatives** (*richtingsafgeleide*), which do measure this rate of change.

We begin with a definition.

Definition 16.15 (Directional derivative)
Let $z = f(x,y)$ be continuous on a set S and let $\vec{u} = (u_1,u_2)$ be a unit vector. For all points (x,y), the **directional derivative of f at (x,y) in the direction of $\vec{u}$** is

$$D_{\vec{u}}f(x,y) = \lim_{h\to 0}\frac{f(x+hu_1, y+hu_2) - f(x,y)}{h}.$$

The partial derivatives f_x and f_y are defined with similar limits, but only x or y varies with h, not both. Here both x and y vary with a weighted h, determined by a particular unit vector $\vec{u}$. In practice it can be a very difficult limit to evaluate, so we need an easier way of taking directional derivatives.

For that purpose, let us define a new function of a single variable,

$$g(z) = f(x_0 + az, y_0 + bz),$$

where x_0, y_0, a, and b are some fixed numbers. Then, by the definition of the derivative for functions of a single variable we have,

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

while the derivative at $z=0$ is given by,

$$g'(0) = \lim_{h\to 0} \frac{g(h)-g(0)}{h}.$$

If we now substitute our expression for $g(z)$ we get,

$$g'(0) = \lim_{h\to 0} \frac{g(h)-g(0)}{h} = \lim_{h\to 0} \frac{f(x_0+ah, y_0+bh)-f(x_0,y_0)}{h} = D_{\vec{\mathbf{u}}} f(x_0,y_0). \tag{16.7}$$

Now, let us look at this from another perspective and rewrite $g(z)$ as follows,

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

where $x = x_0 + az$ and $y = y_0 + bz$. We can now use the chain rule to compute,

$$g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = f_x(x,y)a + f_y(x,y)b. \tag{16.8}$$

If we now take $z=0$ we will get that $x = x_0$ and $y = y_0$ and plug these into Equation (16.8) we get

$$g'(0) = f_x(x_0,y_0)a + f_y(x_0,y_0)b. \tag{16.9}$$

Now, simply equate Equations (16.7) and (16.9) to get that

$$D_{\vec{\mathbf{u}}} f(x_0,y_0) = f_x(x_0,y_0)a + f_y(x_0,y_0)b.$$

If we now go back to allowing x and y to be any number we get the following formula for computing directional derivatives:

$$D_{\vec{\mathbf{u}}} f(x,y) = f_x(x,y)a + f_y(x,y)b.$$

This leads to the following theorem.

Theorem 16.10 (Directional derivatives)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0), and let $\vec{\mathbf{u}} = (u_1,u_2)$ be a unit vector. The directional derivative of f at (x_0,y_0) in the direction of $\vec{\mathbf{u}}$ is

$$D_{\vec{\mathbf{u}}} f(x_0,y_0) = \big(f_x, f_y\big) \cdot (u_1,u_2) = f_x(x_0,y_0)u_1 + f_y(x_0,y_0)u_2.$$

Example 16.23

Let $z = 14 - x^2 - y^2$ and let $P = (1,2)$. Find the directional derivative of f, at P, in the following directions:

1. toward the point $Q = (3,4)$,
2. in the direction of $(2,-1)$, and
3. toward the origin.

The surface is plotted in Figure 16.12, where the point $P = (1,2)$ is indicated in the xy-plane as well as the point $(1,2,9)$ which lies on the surface of f.

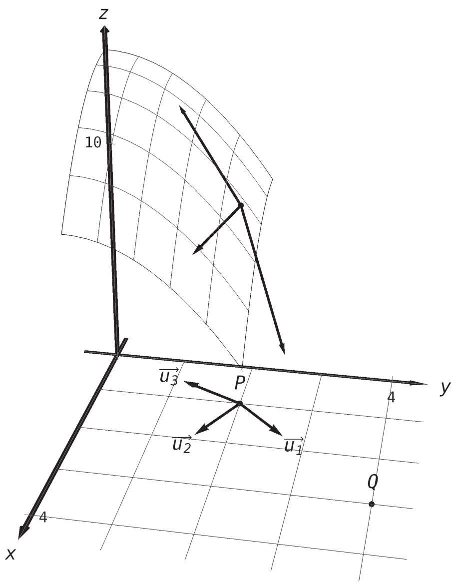


Figure 16.12: Understanding the directional derivative in Example 16.23.

Solution

We find that $f_x(x,y) = -2x$ and $f_x(1,2) = -2$; $f_y(x,y) = -2y$ and $f_y(1,2) = -4$.

1. Let $\vec{\mathbf{u}}_1$ be the unit vector that points from the point $P = (1,2)$ to the point $Q = (3,4)$, as shown in the figure. The vector $\overrightarrow{PQ} = \langle 2,2\rangle$; the unit vector in this direction is $\vec{\mathbf{u}}_1 = \left\langle 1/\sqrt{2}, 1/\sqrt{2}\right\rangle$. Thus the directional derivative of f at $(1,2)$ in the direction of $\vec{\mathbf{u}}_1$ is

$$D_{\vec{\mathbf{u}}_1} f(1,2) = -2\left(\frac{1}{\sqrt{2}}\right) + (-4)\left(\frac{1}{\sqrt{2}}\right) = -\frac{6}{\sqrt{2}} \approx -4.24.$$

   Thus the instantaneous rate of change in moving from the point $(1,2,9)$ on the surface in the direction of $\vec{\mathbf{u}}_1$ (which points toward the point Q) is about -4.24. Moving in this direction moves one steeply downward.

2. We seek the directional derivative in the direction of $\langle 2,-1\rangle$. The unit vector in this direction is $\vec{\mathbf{u}}_2 = \left\langle 2/\sqrt{5}, -1/\sqrt{5}\right\rangle$. Thus the directional derivative of f at $(1,2)$ in the direction of $\vec{\mathbf{u}}_2$ is

$$D_{\vec{\mathbf{u}}_2} f(1,2) = -2\left(\frac{2}{\sqrt{5}}\right) + (-4)\left(-\frac{1}{\sqrt{5}}\right) = 0.$$

   Starting on the surface of f at $(1,2)$ and moving in the direction of $\langle 2,-1\rangle$ (or $\vec{\mathbf{u}}_2$) results in no instantaneous change in z-value. This is analogous to standing on the side of a hill and choosing a direction to walk that does not change the elevation. One neither walks up nor down, rather just along the side of the hill.

3. At $P = (1,2)$, the direction towards the origin is given by the vector $\langle -1,-2\rangle$; the unit vector in this direction is $\vec{\mathbf{u}}_3 = \left\langle -1/\sqrt{5}, -2/\sqrt{5}\right\rangle$. The directional derivative of f at P in the direction

of the origin is

$$D_{\vec{\boldsymbol{u}}_3}f(1,2) = -2\left(-\frac{1}{\sqrt{5}}\right) + (-4)\left(-\frac{2}{\sqrt{5}}\right) = \frac{10}{\sqrt{5}} \approx 4.47.$$

Moving towards the origin means walking uphill quite steeply, with an initial slope of about 4.47.

16.6.2 The gradient

As we study directional derivatives, it will help to make an important connection between the unit vector $\vec{\boldsymbol{u}} = (u_1, u_2)$ that describes the direction and the partial derivatives f_x and f_y. We start with a definition.

Definition 16.16 (Gradient)
Let $z = f(x,y)$ be differentiable on a set S that contains the point (x_0, y_0).

1. The **gradient** (*gradiënt*) of f is

$$\nabla f(x,y) = \big(f_x(x,y), f_y(x,y)\big).$$

2. The **gradient** of f at (x_0, y_0) is

$$\nabla f(x_0,y_0) = \big(f_x(x_0,y_0), f_y(x_0,y_0)\big).$$

The symbol ∇ is named **nabla**, derived from the Greek name of a Jewish harp. Oddly enough, in mathematics the expression ∇f is pronounced del f.

To simplify notation, we often express the gradient as $\nabla f = \big(f_x, f_y\big)$. The gradient allows us to compute directional derivatives in terms of a dot product:

$$D_{\vec{\boldsymbol{u}}}f = \nabla f \cdot \vec{\boldsymbol{u}}. \tag{16.10}$$

The properties of the dot product studied in Chapter 6 allow us to investigate the properties of the directional derivative. Given that the directional derivative gives the instantaneous rate of change of z when moving in the direction of $\vec{\boldsymbol{u}}$, three questions naturally arise:

1. In what direction(s) is the change in z the greatest (i.e., the steepest uphill)?
2. In what direction(s) is the change in z the least (i.e., the steepest downhill)?
3. In what direction(s) is there no change in z?

Relying on the geometric interpretation of the dot product (Theorem 6.2), we have

$$\nabla f \cdot \vec{\boldsymbol{u}} = \big\|\nabla f\big\| \, \|\vec{\boldsymbol{u}}\| \cos(\theta) = \big\|\nabla f\big\| \cos(\theta), \tag{16.11}$$

where θ is the angle between the gradient and $\vec{\boldsymbol{u}}$. (Since $\vec{\boldsymbol{u}}$ is a unit vector, $\|\vec{\boldsymbol{u}}\| = 1$.) This equation allows us to answer the three questions stated previously.

1. Equation (16.11) is maximized when $\cos(\theta) = 1$, i.e., when the gradient and $\vec{\boldsymbol{u}}$ have the same direction. We conclude the gradient points in the direction of greatest z change.

2. Equation (16.11) is minimized when $\cos(\theta) = -1$, i.e., when the gradient and $\vec{u}$ have opposite directions. We conclude the gradient points in the opposite direction of the least z change.

3. Equation (16.11) is 0 when $\cos(\theta) = 0$, i.e., when the gradient and $\vec{u}$ are orthogonal to each other. We conclude the gradient is orthogonal to directions of no z change.

This result is rather amazing. Once again imagine standing in a rolling meadow and face the direction that leads you steepest uphill. Then the direction that leads steepest downhill is directly behind you, and side–stepping either left or right (i.e., moving perpendicularly to the direction you face) does not change your elevation at all.

Recall that a level curve is defined as a curve in the xy-plane along which the z-values of a function do not change. Let a surface $z = f(x,y)$ be given, and let us represent one such level curve as a vector–valued function, $\vec{r}(t) = \big(x(t), y(t)\big)$. As the output of f does not change along this curve, $f\big(x(t), y(t)\big) = c$ for all t, for some constant c.

Since f is constant for all t, $\frac{df}{dt} = 0$. By the multivariable chain rule, we also know

$$\begin{aligned}\frac{df}{dt} &= f_x(x,y)x'(t) + f_y(x,y)y'(t)\\ &= \big(f_x(x,y), f_y(x,y)\big)\cdot\big(x'(t), y'(t)\big)\\ &= \nabla f\cdot\vec{r}\,'(t)\\ &= 0.\end{aligned}$$

This last equality states $\nabla f\cdot\vec{r}\,'(t) = 0$: the gradient is orthogonal to the derivative of $\vec{r}$, meaning the gradient is orthogonal to the graph of $\vec{r}$. Our conclusion: at any point on a surface, the gradient at that point is orthogonal to the level curve that passes through that point.

We restate these ideas in a theorem, then use them in an example.

Theorem 16.11 (The gradient and directional derivatives)
Let $z = f(x,y)$ be differentiable on a set S with gradient ∇f, let $P = (x_0, y_0)$ be a point in S and let $\vec{u}$ be a unit vector.

1. *The maximum value of $D_{\vec{u}}f(x_0,y_0)$ is $\big\|\nabla f(x_0,y_0)\big\|$; the direction of maximal z increase is $\nabla f(x_0,y_0)$.*

2. *The minimum value of $D_{\vec{u}}f(x_0,y_0)$ is $-\big\|\nabla f(x_0,y_0)\big\|$; the direction of maximal z decrease is $-\nabla f(x_0,y_0)$.*

3. *At P, $\nabla f(x_0,y_0)$ is orthogonal to the level curve passing through $\big(x_0, y_0, f(x_0,y_0)\big)$.*

We now illustrate how to find the directions of maximal increase and decrease.

Example 16.24

Let $f(x,y) = \sin(x)\cos(y)$ and let $P = (\pi/3, \pi/3)$. Find the directions of maximal increase and decrease, and find a direction where the instantaneous rate of z change is 0.

Solution

We begin by finding the gradient. We have that $f_x = \cos(x)\cos(y)$ and $f_y = -\sin(x)\sin(y)$, thus

$$\nabla f = \big(\cos(x)\cos(y), -\sin(x)\sin(y)\big)$$

and, at P

$$\nabla f\left(\frac{\pi}{3},\frac{\pi}{3}\right)=\left(\frac{1}{4},-\frac{3}{4}\right).$$

Thus the direction of maximal increase is $\langle 1/4,-3/4\rangle$. In this direction, the instantaneous rate of z change is $\|\langle 1/4,-3/4\rangle\|=\sqrt{10}/4\approx 0.79$.

Figure 16.13 shows the surface. The gradient is drawn at P with a dashed line (because of the nature of this surface, the gradient points into the surface). Let $\vec{u}=\langle u_1,u_2\rangle$ be the unit vector in the direction of ∇f at P. The graph also contains the vector $\langle u_1,u_2,\|\nabla f\|\rangle$. This vector has a run of 1 (because in the xy-plane it moves 1 unit) and a rise of $\|\nabla f\|$, hence we can think of it as a vector with slope of $\|\nabla f\|$ in the direction of ∇f, helping us visualize how steep the surface is in its steepest direction.

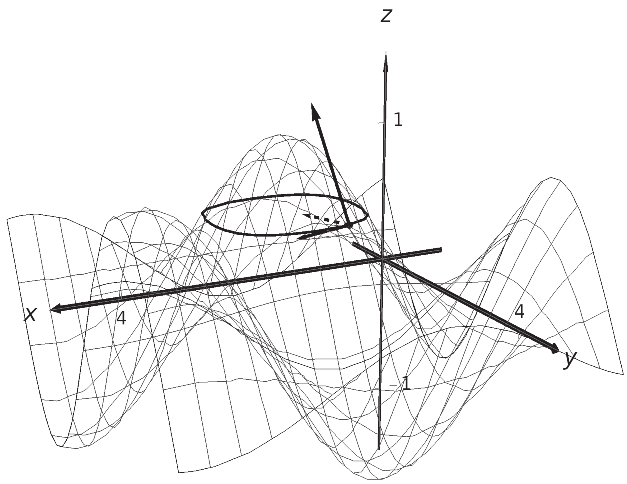


Figure 16.13: Graphing the surface and important directions in Example 16.24.

The direction of maximal decrease is $\langle -1/4,3/4\rangle$; in this direction the instantaneous rate of z change is $-\sqrt{10}/4\approx -0.79$.

Any direction orthogonal to ∇f is a direction of no z change. We have two choices: the direction of $\langle 3,1\rangle$ and the direction of $\langle -3,-1\rangle$. The unit vector in the direction of $\langle 3,1\rangle$ is shown in the graph as well. The level curve at $z=\sqrt{3}/4$ is drawn: recall that along this curve the z-values do not change. Since $\langle 3,1\rangle$ is a direction of no z-change, this vector is tangent to the level curve at P.

It is as well important to figure out when $\nabla f=0$.

Example 16.25

Let $f(x,y)=-x^2+2x-y^2+2y+1$. Find the directional derivative of f in any direction at $P=(1,1)$.

Solution

We find $\nabla f=\langle -2x+2,-2y+2\rangle$. At P, we have $\nabla f(1,1)=\langle 0,0\rangle$. According to Theorem 16.11, this is the direction of maximal increase. However, $\langle 0,0\rangle$ is directionless; it has no displacement. And regardless of the unit vector $\vec{u}$ chosen, $D_{\vec{u}}f=0$.

Figure 16.14 helps us understand what this means. We can see that P lies at the top of a paraboloid. In all directions, the instantaneous rate of change is 0. So what is the direction of maximal increase? It is fine to give an answer of $\vec{0}=\langle 0,0\rangle$, as this indicates that all directional derivatives are 0.

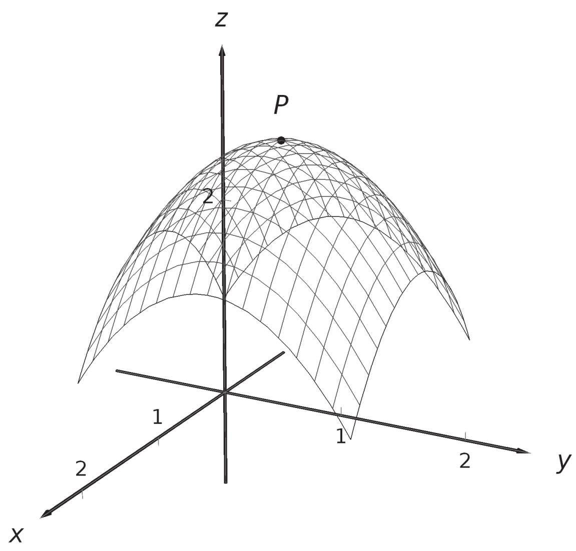


Figure 16.14: At the top of a paraboloid, all directional derivatives are 0.

In Mathematica, we could have computed the gradient in Example 16.25 using the command `Grad` as follows

```
In[30]:= Grad[-x^2 + 2*x - y^2 + 2*y + 1, {x, y}]

Out[30]= {2-2 x,2-2 y}
```

The fact that the gradient of a surface always points in the direction of steepest increase/decrease is very useful, as illustrated in the following example.

Example 16.26

Consider the surface given by $f(x,y) = 20 - x^2 - 2y^2$. Water is poured on the surface at $(1, 1/4)$. What path does it take as it flows downhill?

Solution

Let $\vec{r}(t) = \big(x(t), y(t)\big)$ be the vector–valued function describing the path of the water in the xy-plane; we seek $x(t)$ and $y(t)$. We know that water will always flow downhill in the steepest direction; therefore, at any point on its path, it will be moving in the direction of $-\nabla f$. We ignore the physical effects of momentum on the water. Thus $\vec{r}\,'(t)$ will be parallel to ∇f, and there is some constant c such that $c\nabla f = \vec{r}\,'(t) = \big(x'(t), y'(t)\big)$.

We find $\nabla f = (-2x, -4y)$ and write $x'(t)$ as $\frac{dx}{dt}$ and $y'(t)$ as $\frac{dy}{dt}$. Then

$$\begin{aligned} c\nabla f &= \big(x'(t), y'(t)\big) \\ \Leftrightarrow \qquad (-2cx, -4cy) &= \left(\frac{dx}{dt}, \frac{dy}{dt}\right). \end{aligned}$$

This implies

$$-2cx = \frac{dx}{dt} \qquad \text{and} \qquad -4cy = \frac{dy}{dt},$$

i.e.

$$c = -\frac{1}{2x}\frac{dx}{dt} \quad \text{and} \quad c = -\frac{1}{4y}\frac{dy}{dt}.$$

As c equals both expressions, we have

$$\frac{1}{2x}\frac{dx}{dt} = \frac{1}{4y}\frac{dy}{dt}.$$

To find an explicit relationship between x and y, we can integrate both sides with respect to t. Recall from our study of differentials that $\frac{dx}{dt}dt = dx$. Thus:

$$\begin{aligned} & \int \frac{1}{2x}\frac{dx}{dt}\,dt = \int \frac{1}{4y}\frac{dy}{dt}\,dt \\ \Leftrightarrow \quad & \int \frac{1}{2x}\,dx = \int \frac{1}{4y}\,dy \\ \Leftrightarrow \quad & \frac{1}{2}\ln|x| = \frac{1}{4}\ln|y| + C_1 \\ \Leftrightarrow \quad & 2\ln|x| = \ln|y| + 4C_1 \\ \Leftrightarrow \quad & \ln(x^2) = \ln|y| + 4C_1. \end{aligned}$$

Now raise both sides as a power of e:

$$\begin{aligned} & x^2 = e^{\ln|y| + 4C_1} \\ \Leftrightarrow \quad & x^2 = e^{\ln|y|}e^{4C_1}. \end{aligned}$$

From which it follows that

$$x^2 = yC_2,$$

where $C_2 = \pm e^{4C_1}$, or alternatively

$$Cx^2 = y,$$

where $C = 1/C_2$. As the water started at the point $(1, 1/4)$, we can solve for C:

$$C(1)^2 = \frac{1}{4} \quad \Leftrightarrow \quad C = \frac{1}{4}.$$

Thus the water follows the curve $y = x^2/4$ in the xy-plane. The surface and the path of the water is graphed in Figure 16.15(a). In Figure 16.15(b), the level curves of the surface are plotted in the xy-plane, along with the curve $y = x^2/4$. Notice how the path intersects the level curves at right angles. As the path follows the gradient downhill, this reinforces the fact that the gradient is orthogonal to level curves.

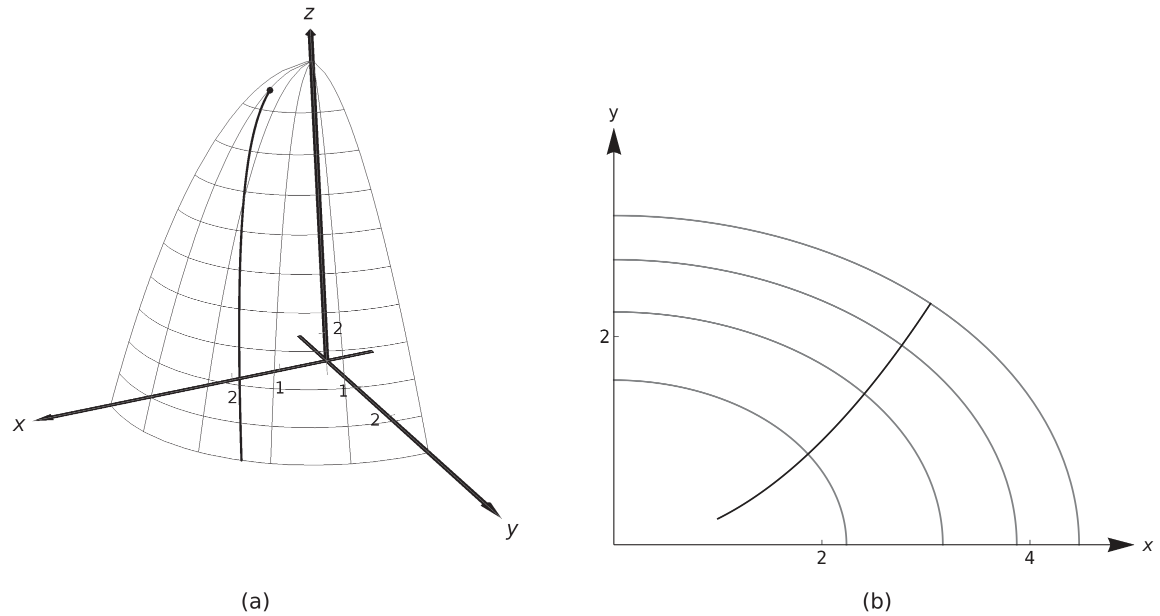


Figure 16.15: A graph of the surface described in Example 16.26 along with the path in the xy-plane with the level curves.

16.6.3 Functions of n variables

The concepts of directional derivatives and the gradient are easily extended to n variables. We combine the concepts behind Definitions 16.15 and 16.16 and Theorem 16.10 into one set of definitions.

Definition 16.17 (Directional derivatives and gradient with n variables)
Let $w = f(\mathbf{x})$ be differentiable on a set D and let $\vec{\mathbf{u}}$ be a unit vector in $\mathbb{R}^n$.

1. The **gradient of** f is $\nabla f = \mathbf{f_x} = \left(f_{x_1}, f_{x_2}, \ldots, f_{x_n}\right)$.
2. The **directional derivative of** f **in the direction of** $\vec{\mathbf{u}}$ is

$$D_{\vec{\mathbf{u}}} f = \nabla f \cdot \vec{\mathbf{u}}.$$

The same properties of the gradient given in Theorem 16.11, hold for a function of three variables.

Theorem 16.12 (The gradient and directional derivatives with n variables)
Let $w = f(\mathbf{x})$ be differentiable on a set D, let ∇f be the gradient of f, and let $\vec{\mathbf{u}}$ be a unit vector.

1. *The maximum value of $D_{\vec{\mathbf{u}}} f$ is $\left\|\nabla f\right\|$, obtained when the angle between ∇f and $\vec{\mathbf{u}}$ is 0, i.e., the direction of maximal increase is ∇f.*
2. *The minimum value of $D_{\vec{\mathbf{u}}} f$ is $-\left\|\nabla f\right\|$, obtained when the angle between ∇f and $\vec{\mathbf{u}}$ is π, i.e., the direction of maximal decrease is $-\nabla f$.*
3. *$D_{\vec{\mathbf{u}}} f = 0$ when ∇f and $\vec{\mathbf{u}}$ are orthogonal.*

We interpret the third statement of the theorem as the gradient is orthogonal to level surfaces, the three–variable analogue to level curves.

Example 16.27

If a point source S is radiating energy, the intensity I at a given point P in space is inversely proportional to the square of the distance between S and P. That is, when $S = (0,0,0)$, it holds that

$$I(x,y,z) = \frac{k}{x^2+y^2+z^2}$$

for some constant k.

Let $k = 1$, let $\vec{\boldsymbol{u}} = (2/3, 2/3, 1/3)$ be a unit vector, and let $P = (2,5,3)$. Measure distances in centimetres. Find the directional derivative of I at P in the direction of $\vec{\boldsymbol{u}}$, and find the direction of greatest intensity increase at P.

Solution

We need the gradient ∇I, meaning we need I_x, I_y and I_z. Each partial derivative requires a simple application of the quotient rule, giving

$$\begin{aligned} \nabla I &= \left(\frac{-2x}{(x^2+y^2+z^2)^2}, \frac{-2y}{(x^2+y^2+z^2)^2}, \frac{-2z}{(x^2+y^2+z^2)^2}\right) \\ \Rightarrow \quad \nabla I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\right) \approx (-0.003, -0.007, -0.004) \\ \Rightarrow \quad D_{\vec{\boldsymbol{u}}} I &= \nabla I(2,5,3) \cdot \vec{\boldsymbol{u}} \\ &= -\frac{17}{2166} \approx -0.0078. \end{aligned}$$

The directional derivative tells us that moving in the direction of $\vec{\boldsymbol{u}}$ from P results in a decrease in intensity of about -0.008 units per centimetre. The intensity is decreasing as $\vec{\boldsymbol{u}}$ moves one farther from the origin than P.

The gradient gives the direction of greatest intensity increase. Notice that

$$\begin{aligned} \nabla I(2,5,3) &= \left(\frac{-4}{1444}, \frac{-10}{1444}, \frac{-6}{1444}\right) \\ &= \frac{2}{1444}(-2,-5,-3). \end{aligned}$$

That is, the gradient at $(2,5,3)$ is pointing in the direction of $(-2,-5,-3)$, that is, towards the origin. That should make intuitive sense: the greatest increase in intensity is found by moving towards to source of the energy.

The directional derivative allows us to find the instantaneous rate of z change in any direction at a point. We can use these instantaneous rates of change to define lines and planes that are tangent to a surface at a point, which is the topic of the next section.

Using the gradient of functions of n variables introduced in Definition 16.17, we can easily state the multivariable counterpart of the mean value theorem of differentiation (Theorem 10.4).

Theorem 16.13 (Mean value theorem for functions of *n* variables)
Let f be differentiable on D, let $\vec{\boldsymbol{a}}$ and $\vec{\boldsymbol{b}}$ belong to D and let ∇f be the gradient of f, then there exists a point $\vec{\boldsymbol{c}}$ on the line segment L connecting $\vec{\boldsymbol{a}}$ and $\vec{\boldsymbol{b}}$ for which it holds that

$$\nabla f(\vec{\boldsymbol{c}})\cdot\left(\vec{\boldsymbol{b}}-\vec{\boldsymbol{a}}\right)=f(\vec{\boldsymbol{b}})-f(\vec{\boldsymbol{a}}).$$

The key of the proof of this theorem is to design an appropriate function of one variable to which we can then apply the mean value theorem for functions of one variable (Theorem 10.4). For that purpose, we ought to realize that as we let t range from 0 to 1, $\vec{\boldsymbol{a}}+t(\vec{\boldsymbol{b}}-\vec{\boldsymbol{a}})$ traces out the line segment L. So, the idea of the proof is to apply the one-variable mean-value theorem to the function

$$\vec{\boldsymbol{\phi}}(t)=\vec{\boldsymbol{a}}+t\left(\vec{\boldsymbol{b}}-\vec{\boldsymbol{a}}\right)=(1-t)\vec{\boldsymbol{a}}+t\vec{\boldsymbol{b}},$$

for $t\in[0,1]$. Clearly, $\vec{\boldsymbol{\phi}}(t)$ is differentiable with $\vec{\boldsymbol{\phi}}'(t)=\vec{\boldsymbol{b}}-\vec{\boldsymbol{a}}$. Consequently, the composition $g=f\circ\vec{\boldsymbol{\phi}}$ is differentiable by the chain rule. Since g provides a mapping from the unit interval to the real numbers, i.e. $g:[0,1]\mapsto\mathbb{R}$, the mean value theorem for functions of one variable applied to g gives us a point ξ in $[0,1]$, such that

$$g(1)-g(0)=g'(\xi).$$

Now, let $\mathbf{c}=\vec{\boldsymbol{\phi}}(\xi)\in L$. We then have

$$\begin{aligned}
f(\vec{\boldsymbol{b}})-f(\vec{\boldsymbol{a}}) &= g(1)-g(0)\\
&= g'(\xi),\\
&= \nabla f\left(\vec{\boldsymbol{\phi}}(\xi)\right)\cdot\vec{\boldsymbol{\phi}}'(\xi)\\
&= \nabla f(\vec{\boldsymbol{c}})\cdot(\vec{\boldsymbol{b}}-\vec{\boldsymbol{a}}),
\end{aligned}$$

which concludes the proof. For comprehensiveness, it should be noted that the transition from $g'(\xi)$ to $\nabla f\left(\vec{\boldsymbol{\phi}}(\xi)\right)\cdot\vec{\boldsymbol{\phi}}'(\xi)$ in the right-hand side follows from clever use of the chain rule. More precisely, f is a function of n variables x_i applied to the output of the single-variable vector-valued function $\vec{\boldsymbol{\phi}}(t)$ with components $x_i=\phi_i(t)$, so the chain rule (Theorem 16.6) immediately gives

$$\frac{dg}{dt}=\frac{\partial f}{\partial x_1}\frac{dx_1}{dt}+\frac{\partial f}{\partial x_2}\frac{dx_2}{dt}+\cdots+\frac{\partial f}{\partial x_n}\frac{dx_n}{dt},$$

which can be written using the dot product as

$$\frac{dg}{dt}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\ldots,\frac{\partial f}{\partial x_n}\right)\cdot\left(\frac{dx_1}{dt},\frac{dx_2}{dt},\ldots,\frac{dx_n}{dt}\right).$$

16.7 Tangent lines, normal lines, and tangent planes

16.7.1 Tangent and normal lines

Derivatives and tangent lines go hand–in–hand. Given $y=f(x)$, the line tangent to the graph of f at $x=x_0$ is the line through $\left(x_0,f(x_0)\right)$ with slope $f'(x_0)$; that is, the slope of the tangent line is the instantaneous rate of change of f at x_0. When dealing with functions of two variables, the graph is no longer a curve but a surface. At a given point on the surface, it seems there are many lines that fit our intuition of being tangent to the surface.

In Figure 16.16 we see lines that are tangent to curves in space. Since each curve lies on a surface, it makes sense to say that the lines are also tangent to the surface. The next definition formally defines what it means to be tangent to a surface.

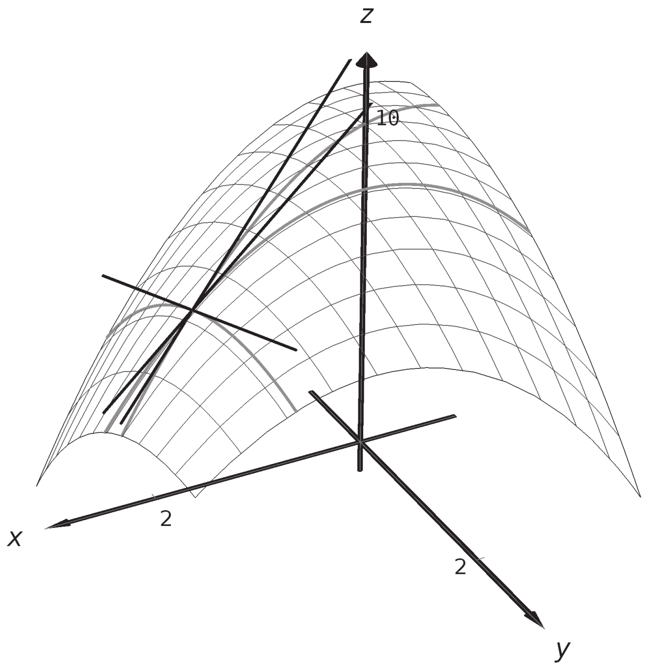


Figure 16.16: Showing various lines tangent to a surface.

Definition 16.18 (Directional tangent line)
Let $z = f(x, y)$ be differentiable on a set S containing (x_0, y_0) and let $\vec{u} = \langle u_1, u_2 \rangle$ be a unit vector.

1. The line ℓ_x through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\langle 1, 0, f_x(x_0, y_0)\rangle$ is **the tangent line to f in the direction of x at** (x_0, y_0).
2. The line ℓ_y through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\langle 0, 1, f_y(x_0, y_0)\rangle$ is **the tangent line to f in the direction of y at** (x_0, y_0).
3. The line $\ell_{\vec{u}}$ through $\big(x_0, y_0, f(x_0, y_0)\big)$ parallel to $\langle u_1, u_2, D_{\vec{u}} f(x_0, y_0)\rangle$ is **the tangent line to f in the direction of $\vec{u}$ at** (x_0, y_0).

It is instructive to consider each of three directions given in the definition in terms of slope. The direction of ℓ_x is $\langle 1, 0, f_x(x_0, y_0)\rangle$; that is, the "run" is one unit in the x-direction and the rise is $f_x(x_0, y_0)$ units in the z-direction. Note how the slope is just the partial derivative with respect to x. A similar statement can be made for ℓ_y. The direction of $\ell_{\vec{u}}$ is $\langle u_1, u_2, D_{\vec{u}} f(x_0, y_0)\rangle$; the run is one unit in the $\vec{u}$ direction (where $\vec{u}$ is a unit vector) and the rise is the directional derivative of z in that direction.

Definition 16.18 leads to the following parametric equations of directional tangent lines:

$$\ell_x(t) = \begin{cases} x &= x_0 + t \\ y &= y_0 \\ z &= z_0 + f_x(x_0, y_0)t, \end{cases} \quad \ell_y(t) = \begin{cases} x &= x_0 \\ y &= y_0 + t \\ z &= z_0 + f_y(x_0, y_0)t \end{cases} \quad \text{and } \ell_{\vec{u}}(t) = \begin{cases} x &= x_0 + u_1 t \\ y &= y_0 + u_2 t \\ z &= z_0 + D_{\vec{u}} f(x_0, y_0)t. \end{cases}$$

where $z_0 = f(x_0, y_0)$.

Example 16.28

Find the lines tangent to the surface $z = \sin(x)\cos(y)$ at $(\pi/2, \pi/2)$ in the x- and y- directions and also in the direction of $\vec{v} = \langle -1, 1 \rangle$.

Solution

The partial derivatives with respect to x and y are:

$$f_x(x,y) = \cos(x)\cos(y)$$
$$f_y(x,y) = -\sin(x)\sin(y).$$

from which it follows that $f_x(\pi/2, \pi/2) = 0$ and $f_y(\pi/2, \pi/2) = -1$. At $(\pi/2, \pi/2)$, the z-value is 0.

Thus the parametric equations of the line tangent to f at $(\pi/2, \pi/2)$ in the directions of x and y are:

$$\ell_x(t) = \begin{cases} x = \pi/2 + t \\ y = \pi/2 \\ z = 0 \end{cases} \quad \text{and} \quad \ell_y(t) = \begin{cases} x = \pi/2 \\ y = \pi/2 + t \\ z = -t. \end{cases}$$

The two lines are shown with the surface in Figure 16.17(a).

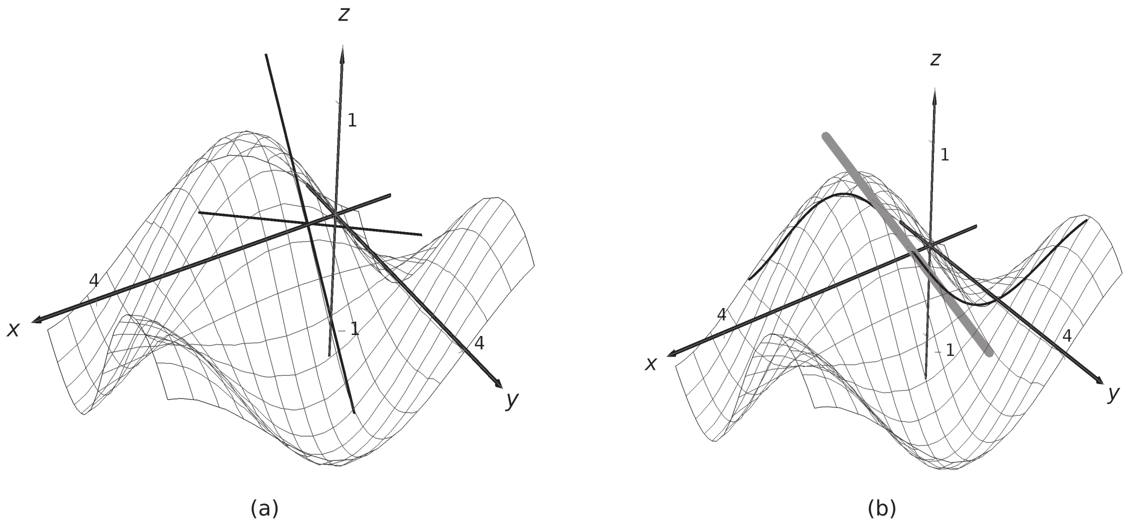


Figure 16.17: A surface and directional tangent lines in Example 16.28.

To find the equation of the tangent line in the direction of $\vec{v}$, we first find the unit vector in the direction of $\vec{v}$: $\vec{u} = \left\langle -1/\sqrt{2}, 1/\sqrt{2} \right\rangle$. The directional derivative at $(\pi/2, \pi, 2)$ in the direction of $\vec{u}$ is

$$D_{\vec{u}} f\left(\frac{\pi}{2}, \pi, 2\right) = \langle 0, -1 \rangle \cdot \left\langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = -\frac{1}{\sqrt{2}}.$$

Thus the directional tangent line is

$$\ell_{\vec{u}}(t) = \begin{cases} x = \dfrac{\pi}{2} - \dfrac{t}{\sqrt{2}} \\ y = \dfrac{\pi}{2} + \dfrac{t}{\sqrt{2}} \\ z = -\dfrac{t}{\sqrt{2}} \end{cases}.$$

The curve through $(\pi/2, \pi/2, 0)$ in the direction of $\vec{\mathbf{v}}$ is shown in Figure 16.17(b) along with $\ell_{\vec{\mathbf{u}}}(t)$.

The following example shows that the points on surfaces where all tangent lines have a slope of 0 can give us some information about the extrema of functions of several variables.

Example 16.29

Let $f(x,y) = 4xy - x^4 - y^4$. Find the equations of all directional tangent lines to f at $(1,1)$.

Solution

First note that $f(1,1) = 2$. We need to compute directional derivatives, so we need ∇f. We begin by computing partial derivatives.

$$f_x = 4y - 4x^3 \quad \Rightarrow \quad f_x(1,1) = 0; \qquad f_y = 4x - 4y^3 \quad \Rightarrow \quad f_y(1,1) = 0.$$

Thus $\nabla f(1,1) = (0,0)$. Let $\vec{\mathbf{u}} = (u_1, u_2)$ be any unit vector. The directional derivative of f at $(1,1)$ will be $D_{\vec{\mathbf{u}}} f(1,1) = (0,0) \cdot (u_1, u_2) = 0$. It does not matter what direction we choose; the directional derivative is always 0. Therefore

$$\ell_{\vec{\mathbf{u}}}(t) = \begin{cases} x = 1 + u_1 t \\ y = 1 + u_2 t \\ z = 2. \end{cases}$$

Figure 16.18 shows a graph of f and the point $(1,1,2)$. Note that this point comes at the top of a hill, and therefore every tangent line through this point will have a slope of 0.

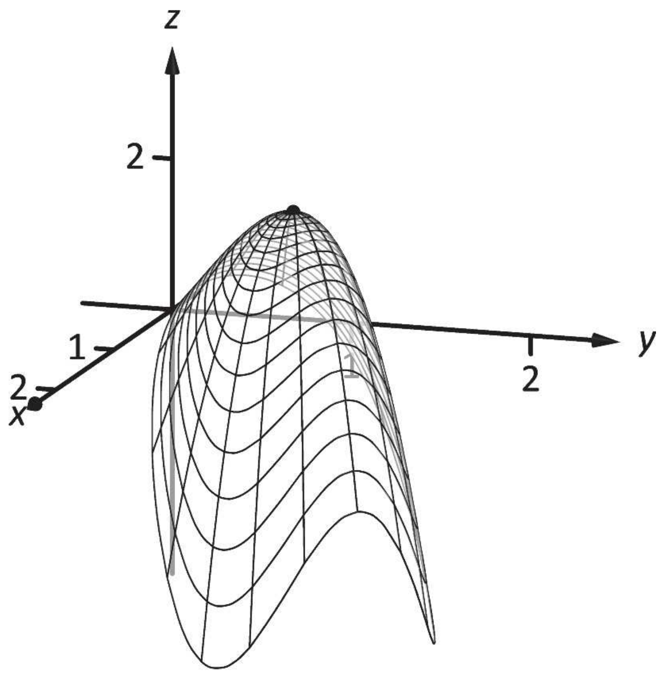


Figure 16.18: Graphing f in Example 16.29.

That is, consider any curve on the surface that goes through this point. Each curve will have a relative maximum at this point, hence its tangent line will have a slope of 0. The following section investigates the points on surfaces where all tangent lines have a slope of 0.

When dealing with a function $y = f(x)$ of one variable, we stated that a line through $(c, f(c))$ was tangent to f if the line had a slope of $f'(c)$ and was normal to f if it had a slope of $-1/f'(c)$. We extend the concept of normal, or orthogonal, to functions of two variables.

Let $z = f(x,y)$ be a differentiable function of two variables. By Definition 16.18, at (x_0,y_0), $\ell_x(t)$ is a line parallel to the vector $\vec{\boldsymbol{d}}_x = \big(1,0,f_x(x_0,y_0)\big)$ and $\ell_y(t)$ is a line parallel to $\vec{\boldsymbol{d}}_y = \big(0,1,f_y(x_0,y_0)\big)$. Since lines in these directions through $\big(x_0,y_0,f(x_0,y_0)\big)$ are tangent to the surface, a line through this point and orthogonal to these directions would be orthogonal, or normal, to the surface. We can use this direction to create a normal line.

The direction of the normal line is orthogonal to $\vec{\boldsymbol{d}}_x$ and $\vec{\boldsymbol{d}}_y$, hence the direction is parallel to $\vec{\boldsymbol{d}}_n = \vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y$. It turns out this cross product has a very simple form:

$$\vec{\boldsymbol{d}}_x \times \vec{\boldsymbol{d}}_y = (1,0,f_x) \times \big(0,1,f_y\big) = \big(-f_x,-f_y,1\big).$$

It is often more convenient to refer to the opposite of this direction, namely $\big(f_x,f_y,-1\big)$. This leads to a definition.

Definition 16.19 (Normal line)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0,y_0) where

$$a = f_x(x_0,y_0) \quad \text{and} \quad b = f_y(x_0,y_0)$$

are defined.

1. A nonzero vector parallel to $\vec{\boldsymbol{n}} = (a,b,-1)$ is orthogonal to f at $P = \big(x_0,y_0,f(x_0,y_0)\big)$.
2. The line ℓ_n through P with direction parallel to $\vec{\boldsymbol{n}}$ is the **normal line** (*normaal*) to f at P.

Thus the parametric equations of the normal line to a surface f at $\big(x_0,y_0,f(x_0,y_0)\big)$ are:

$$\ell_n(t) = \begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = f(x_0,y_0) - t. \end{cases}$$

The direction of the normal line has many uses, one of which is the definition of the **tangent plane** (*raakvlak*) which we define shortly. Another use is in measuring distances from the surface to a point. Given a point Q in space, it is a general geometric concept to define the distance from Q to the surface as being the length of the shortest line segment $\overrightarrow{PQ}$ over all points P on the surface. This, in turn, implies that $\overrightarrow{PQ}$ will be orthogonal to the surface at P. Therefore we can measure the distance from Q to the surface f by finding a point P on the surface such that $\overrightarrow{PQ}$ is parallel to the normal line to f at P.

Example 16.30

Let $f(x,y) = 2 - x^2 - y^2$ and let $Q = (2,2,2)$. Find the distance from Q to the surface defined by f.

Solution

From Definition 16.19, we know that at (x,y) the direction of the normal line will be $\vec{\boldsymbol{d}}_n = (-2x,-2y,-1)$. A point P on the surface will have coordinates $(x,y,2-x^2-y^2)$, so $\overrightarrow{PQ} = \big(2-x,2-y,x^2+y^2\big)$. To find where $\overrightarrow{PQ}$ is parallel to $\vec{\boldsymbol{d}}_n$, we need to find x, y and c such that $c\overrightarrow{PQ} = \vec{\boldsymbol{d}}_n$.

$$\begin{aligned} c\overrightarrow{PQ} &= \vec{\boldsymbol{d}}_n \\ \Rightarrow\quad c\big(2-x,2-y,x^2+y^2\big) &= (-2x,-2y,-1) \end{aligned}$$

This implies

$$\begin{aligned} c(2-x) &= -2x, \\ c(2-y) &= -2y, \\ c(x^2+y^2) &= -1. \end{aligned}$$

In each equation, we can solve for c:

$$c = \frac{-2x}{2-x} = \frac{-2y}{2-y} = \frac{-1}{x^2+y^2}.$$

The first two fractions imply $x=y$, and so the last fraction can be rewritten as $c = -1/(2x^2)$. Then

$$\begin{aligned} && \frac{-2x}{2-x} &= \frac{-1}{2x^2} \\ &\Leftrightarrow & -2x(2x^2) &= -1(2-x) \\ &\Leftrightarrow & 4x^3 &= 2-x \\ &\Leftrightarrow & 4x^3+x-2 &= 0. \end{aligned}$$

This last equation is a cubic, which is not difficult to solve with a numeric solver. We find that $x \approx 0.689$, hence $P = (0.689, 0.689, 1.051)$. We find the distance from Q to the surface of f is

$$\left\|\overrightarrow{PQ}\right\| = \sqrt{(2-0.689)^2+(2-0.689)^2+(2-1.051)^2} = 2.083.$$

We can of course take the concept of measuring the distance from a point to a surface to find a point Q a particular distance from a surface at a given point P on the surface.

16.7.2 Tangent plane

We can use the direction of the normal line to define a plane. With $a = f_x(x_0, y_0)$, $b = f_y(x_0, y_0)$ and $P = \big(x_0, y_0, f(x_0, y_0)\big)$, the vector $\vec{n} = (a, b, -1)$ is orthogonal to f at P. The plane through P with normal vector $\vec{n}$ is therefore tangent to f at P.

Definition 16.20 (Tangent plane)
Let $z = f(x,y)$ be differentiable on a set S containing (x_0, y_0), where $a = f_x(x_0, y_0)$, $b = f_y(x_0, y_0)$, $\vec{n} = (a, b, -1)$ and $P = \big(x_0, y_0, f(x_0, y_0)\big)$.

The plane through P with normal vector $\vec{n}$ is the **tangent plane to** f **at** P (*raakvlak aan f in P*). The standard form of this plane is

$$\begin{aligned} && \vec{n}\cdot\big((x-x_0),(y-y_0),(z-f(x_0,y_0))\big) &= 0 \\ &\Leftrightarrow & a(x-x_0)+b(y-y_0)-\big(z-f(x_0,y_0)\big) &= 0. \end{aligned}$$

Example 16.31

Find the equation of the normal line and tangent plane to $z = -x^2-y^2+2$ at $(0,1)$.

Solution

We find $z_x(x,y) = -2x$ and $z_y(x,y) = -2y$; at $(0,1)$, we have $z_x = 0$ and $z_y = -2$. We take the direction of the normal line, following Definition 16.19, to be $\vec{n} = \langle 0, -2, -1 \rangle$. The line with this direction going through the point $(0,1,1)$ is

$$\ell_n(t) = \begin{cases} x = 0 \\ y = 1 - 2t \\ z = 1 - t \end{cases} \quad \text{or} \quad \ell_n(t) = \langle 0, 1, 1 \rangle + \langle 0, -2, -1 \rangle t.$$

The surface $z = -x^2 - y^2 + 2$, along with the found normal line, is graphed in Figure 16.19(a).

Since we have that $\vec{n} = \langle 0, -2, -1 \rangle$ and $P = (0,1,1)$, the equation of the tangent plane is

$$-2(y-1) - (z-1) = 0.$$

The surface $z = -x^2 - y^2 + 2$ and tangent plane are graphed in Figure 16.19(b).

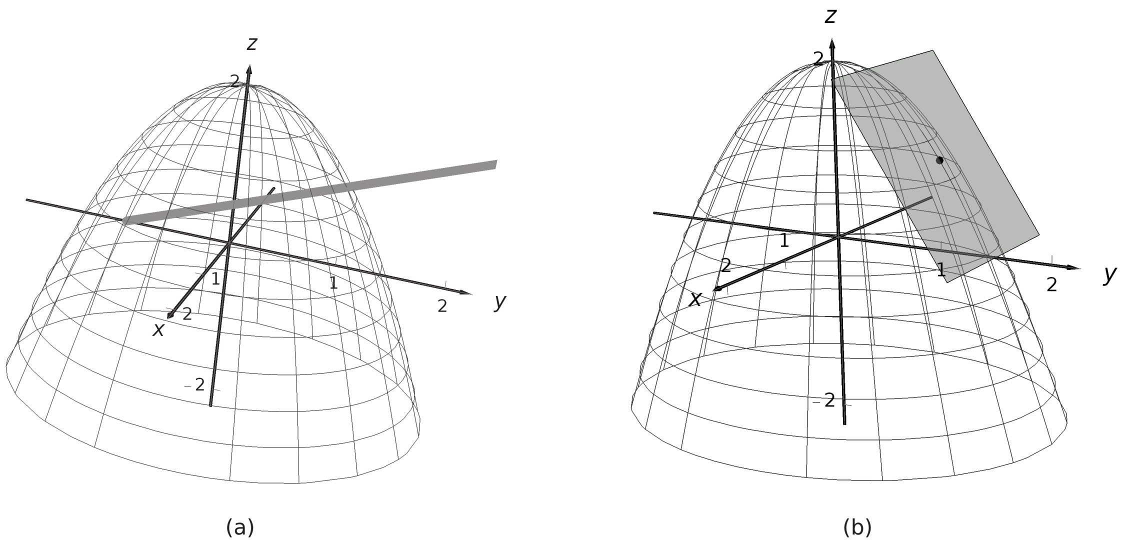


Figure 16.19: Graphing a surface with normal line and tangent plane from Example 16.31.

Just as tangent lines can be used to approximate function values of functions of one variable, tangent planes can be used to achieve this for functions of two variables.

Example 16.32

The point $(3, -1, 4)$ lies on the surface of an unknown differentiable function f where $f_x(3,-1) = 2$ and $f_y(3,-1) = -1/2$. Find the equation of the tangent plane to f at P, and use this to approximate the value of $f(2.9, -0.8)$.

Solution

Knowing the partial derivatives at $(3,-1)$ allows us to form the normal vector to the tangent plane, $\vec{n} = \langle 2, -1/2, -1 \rangle$. Thus the equation of the tangent line to f at P is:

$$2(x-3) - \frac{1}{2}(y+1) - (z-4) = 0 \quad \Leftrightarrow \quad z = 2(x-3) - \frac{1}{2}(y+1) + 4. \tag{16.12}$$

Just as tangent lines provide excellent approximations of curves near their point of intersection, tangent planes provide excellent approximations of surfaces near their point of intersection. So

$f(2.9,-0.8) \approx z(2.9,-0.8) = 3.7$.

This is not a new method of approximation. Compare the right hand expression for z in Equation (16.12) to the total differential:

$$dz = f_x dx + f_y dy \quad \text{and} \quad z = \underbrace{\underbrace{2}_{f_x}\underbrace{(x-3)}_{dx} + \underbrace{-1/2}_{f_y}\underbrace{(y+1)}_{dy}}_{dz} + 4.$$

Thus the new z-value is the sum of the change in z (i.e., dz) and the old z-value (4). As mentioned when studying the total differential, it is not uncommon to know partial derivative information about an unknown function f, and tangent planes are used to give accurate approximations of f.

Recall that when $w = f(x,y,z)$, the gradient $\nabla f = \left(f_x, f_y, f_z\right)$ is orthogonal to level surfaces of f. Given a point (x_0, y_0, z_0), let $c = f(x_0, y_0, z_0)$. Then $f(x,y,z) = c$ is a level surface that contains the point (x_0, y_0, z_0) and $\nabla f(x_0, y_0, z_0)$ is orthogonal to this level surface. So, the gradient at a point gives a vector orthogonal to the surface at that point. This direction can be used to find tangent planes and normal lines.

Example 16.33

Find the equation of the plane tangent to the ellipsoid

$$\frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4} = 1$$

at $P = (1,2,1)$.

Solution

We consider the equation of the ellipsoid as a level surface of a function F of three variables, where $F(x,y,z) = \frac{x^2}{12} + \frac{y^2}{6} + \frac{z^2}{4}$. The gradient is:

$$\begin{aligned} \nabla F(x,y,z) &= \left(F_x, F_y, F_z\right) \\ &= \left(\frac{x}{6}, \frac{y}{3}, \frac{z}{2}\right). \end{aligned}$$

At P, the gradient is $\nabla F(1,2,1) = (1/6, 2/3, 1/2)$. Thus the equation of the plane tangent to the ellipsoid at P is

$$\frac{1}{6}(x-1) + \frac{2}{3}(y-2) + \frac{1}{2}(z-1) = 0.$$

The ellipsoid and tangent plane are graphed in Figure 16.20.

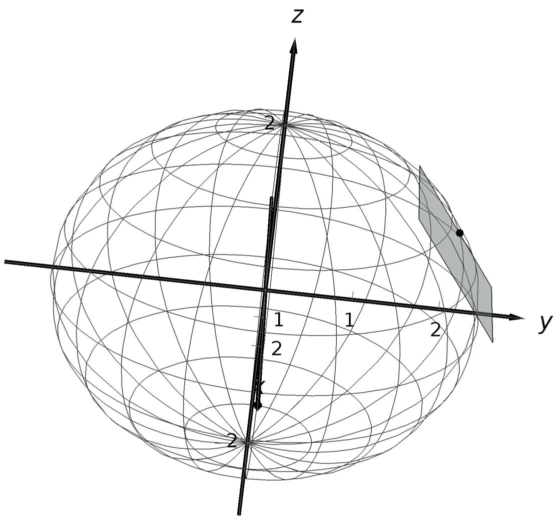


Figure 16.20: An ellipsoid and its tangent plane at a point.

Tangent lines and planes to surfaces have many uses, including the study of instantaneous rates of changes and making approximations. Normal lines also have many uses. In this section we focused on using them to measure distances from a surface. Another interesting application is in computer graphics, where the effects of light on a surface are determined using normal vectors.

The next section investigates another use of partial derivatives: Taylor series expansions of functions of several variables.

16.8 Taylor series expansions

Recall that we found in Section 14.7 a way of rewriting a continuous function of one variable as a series by relying on Taylor's theorem. Having introduced all mathematical tools that are needed to analyse functions of several variables, we are now ready to introduce Taylor's theorem for a function of two variables.

Theorem 16.14 (Taylor's theorem for a function of two variables)
Let f be a C^{n+1} function on a set D containing (x_0,y_0). Then, for each (x,y) in D, there exists (θ_x,θ_y) between (x,y) and (x_0,y_0) such that

$$f(x,y)=\sum_{i=0}^{n}\frac{1}{i!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^i f(x_0,y_0)+R_n(x,y),$$

where

$$f(x,y)=f(x_0,y_0)+\frac{1}{1!}f_x(x_0,y_0)(x-x_0)+\frac{1}{1!}f_y(x_0,y_0)(y-y_0)+\frac{1}{2!}f_{xx}(x_0,y_0)(x-x_0)^2+$$
$$\frac{2}{2!}f_{xy}(x_0,y_0)(x-x_0)(y-y_0)+\frac{1}{2!}f_{yy}(x_0,y_0)(y-y_0)^2+\ldots$$

and the remainder term is given by

$$R_n(x,y) = \frac{1}{(n+1)!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^{n+1} f(\theta_x,\theta_y).$$

For the sake of brevity, we will restrict the proof to functions of two variables, though the same reasoning may be followed to arrive at a full proof of the n variables case.

Essentially, the proof starts off in a similar way as the one of the mean value theorem for functions of n variables (Theorem 16.13) by devising an appropriate one-variable function to which we can apply Taylor's theorem of functions of one variable (Theorem 16.14).

We start by considering a point in the neighbourhood of (x_0, y_0). Let us say, the point $(x_0+\Delta x, y_0+\Delta y)$ and we then look for a formula for $f(x_0+\Delta x, y_0+\Delta y)$. As we will try to use Taylor's theorem for functions of one variable, we consider the vector-valued function $\vec{\boldsymbol{\phi}}: \mathbb{R} \mapsto \mathbb{R}^2$ of one variable:

$$\vec{\boldsymbol{\phi}}(t) = (x_0+t\Delta x, y_0+t\Delta y)$$

for $t \in [0,1]$, which traces out the line segment connecting (x_0, y_0) and $(x_0+\Delta x, y_0+\Delta y)$. Subsequently, we consider the composition

$$g = f \circ \vec{\boldsymbol{\phi}},$$

which is C^{n+1} in agreement with what we concluded in the proof of Theorem 16.13. Consequently, we may apply Taylor's theorem of functions of one variable to g, meaning that for every $t \in [0,1]$ there exists a $\theta_t \in]0,1[$ such that

$$g(t) = g(0) + g'(0)t + \frac{g''(0)}{2!}t^2 + \cdots + \frac{g^{(n)}(0)}{n!}t^n + R_n(t), \tag{16.13}$$

where

$$R_n(t) = \frac{g^{(n+1)}(\theta_t)}{(n+1)!}t^{(n+1)}.$$

Now, note that

$$g(0) = f(x_0, y_0) \qquad \text{and} \qquad g(1) = f(x_0+\Delta x, y_0+\Delta y),$$

while the higher-order derivatives in Equation (16.13) can be obtained using the chain rule. In this way, we obtain

$$\begin{aligned}
g(t) &= f(x_0+t\Delta x, y_0+t\Delta y)\\
g'(t) &= \Delta x\frac{\partial f}{\partial x}(x_0+t\Delta x, y_0+t\Delta y) + \Delta y\frac{\partial f}{\partial y}(x_0+t\Delta x, y_0+t\Delta y)\\
&= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right) f(x_0+t\Delta x, y_0+t\Delta y)\\
g''(t) &= \Delta x^2\frac{\partial^2 f}{\partial x^2}(x_0+t\Delta x, y_0+t\Delta y) + 2\Delta x\Delta y\frac{\partial^2 f}{\partial x\partial y}(x_0+t\Delta x, y_0+t\Delta y) + \Delta y^2\frac{\partial^2 f}{\partial y^2}(x_0+t\Delta x, y_0+t\Delta y)\\
&= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right)^2 f(x_0+t\Delta x, y_0+t\Delta y)\\
&\dots\\
g^{(n)}(t) &= \left(\Delta x\frac{\partial}{\partial x}+\Delta y\frac{\partial}{\partial y}\right)^n f(x_0+t\Delta x, y_0+t\Delta y)
\end{aligned}$$

More specifically, we immediately get

$$g^{(n)}(0) = \left(\Delta x\frac{\partial}{\partial x} + \Delta y\frac{\partial}{\partial y}\right)^n f(x_0, y_0)\,.$$

Consequently, choosing $t = 1$ in Equation (16.13) we arrive at

$$g(1) = f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + \left(\Delta x\frac{\partial}{\partial x} + \Delta y\frac{\partial}{\partial y}\right) f(x_0, y_0) + \cdots$$
$$+ \frac{1}{2!}\left(\Delta x\frac{\partial}{\partial x} + \Delta y\frac{\partial}{\partial y}\right)^2 f(x_0, y_0) + \cdots + \frac{1}{n!}\left(\Delta x\frac{\partial}{\partial x} + \Delta y\frac{\partial}{\partial y}\right)^n f(x_0, y_0) + R_n(x, y)\,,$$

where

$$R_n(x, y) = \frac{1}{(n+1)!}\left(\Delta x\frac{\partial}{\partial x} + \Delta y\frac{\partial}{\partial y}\right)^{n+1} f(x_0 + \theta_t\Delta x, y_0 + \theta_t\Delta y)\,.$$

Finally, acknowledging that $\Delta x = x - x_0$ and $\Delta y = y - y_0$, we can restate the previous expression as

$$g(1) = f(x, y) = f(x_0, y_0) + \left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right) f(x_0, y_0) + \cdots$$
$$+ \frac{1}{2}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^2 f(x_0, y_0) + \cdots + \frac{1}{n!}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^n f(x_0, y_0) + R_n(x, y)\,,$$

$$R_n(x, y) = \frac{1}{(n+1)!}\left((x - x_0)\frac{\partial}{\partial x} + (y - y_0)\frac{\partial}{\partial y}\right)^{n+1} f(\theta_x, \theta_y)\,,$$

where $\theta_x = x_0 + \theta_t(x - x_0)$ and $\theta_y = y_0 + \theta_t(y - y_0)$.

As for functions of one variable, the Talor polynomial of degree n provides the best n-th degree polynomial approximation of $f(x, y)$ near a point (x_0, y_0). For instance, letting $n = 1$, we obtain

$$f(x, y) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Example 16.34

Find a second-degree polynomial approximation to the function

$$f(x, y) = \sqrt{x^2 + y^3}$$

near the point $(1, 2)$ and use it to estimate the value of $\sqrt{1.02^2 + 1.97^3}$.

Solution

For a second-degree approximation we need the values of the partial derivatives of f up to the second order at the point $(1, 2)$. We have

Derivative function			Derivative at $(1,2)$		
$f(x,y)$	$=$	$\sqrt{x^2+y^3}$	$f(1,2)$	$=$	3
$f_x(x,y)$	$=$	$\dfrac{x}{\sqrt{x^2+y^3}}$	$f_x(1,2)$	$=$	$\dfrac{1}{3}$
$f_y(x,y)$	$=$	$\dfrac{3y^2}{2\sqrt{x^2+y^3}}$	$f_x(1,2)$	$=$	2
$f_{xx}(x,y)$	$=$	$\dfrac{y^3}{\left(x^2+y^3\right)^{3/2}}$	$f_{xx}(1,2)$	$=$	$\dfrac{8}{27}$
$f_{xy}(x,y)$	$=$	$\dfrac{-3xy^2}{2\left(x^2+y^3\right)^{3/2}}$	$f_{xy}(1,2)$	$=$	$-\dfrac{2}{9}$
$f_{yy}(x,y)$	$=$	$\dfrac{12x^2y+3y^4}{4\left(x^2+y^3\right)^{3/2}}$	$f_{yy}(1,2)$	$=$	$\dfrac{2}{3}.$

Thus, we get after evaluating the partial derivatives in $(1,2)$

$$f(x,y) \approx 3+\frac{1}{3}(x-1)+2(y-2)+\frac{4}{27}(x-1)^2-\frac{2}{9}(x-1)(y-2)+\frac{1}{3}(y-2)^2.$$

This is the required second-degree Taylor polynomial for f near $(1,2)$. Therefore,

$$\begin{aligned}
\sqrt{1.02^2+1.97^3} &= f(1+0.02, 2-0.03) \\
&\approx 3+\frac{1}{3}(0.02)+2(-0.03)+\frac{4}{27}(0.02)^2-\frac{2}{9}(0.02)(-0.03)+\frac{1}{3}(-0.03)^2 \\
&\approx 2.9471593.
\end{aligned}$$

It can be verified that the true value is 2.9471636, so our approximation is accurate to six significant figures.

Clearly, in line with what we devised for functions of one variable, the Taylor series expansion of a function of two variables is given by

$$f(x,y) = \sum_{i=0}^{+\infty} \frac{1}{i!}\left(\frac{\partial}{\partial x}(x-x_0)+\frac{\partial}{\partial y}(y-y_0)\right)^i f(x_0,y_0),$$

and likewise a Maclaurin series expansion can be formulated.

Example 16.35

Find a second-order Taylor series expansion of the function

$$f(x,y) = e^x \ln(1+y),$$

around the point $(0,0)$.

Solution

In order to compute a second-order Taylor series expansion we first compute the necessary partial derivatives and evaluate these derivatives at the origin:

Derivative function	Derivative at $(0,0)$
$f_x(x,y) = e^x \ln(1+y)$	$f_x(0,0) = 0$
$f_y(x,y) = \dfrac{e^x}{1+y}$	$f_y(0,0) = 1$
$f_{xx}(x,y) = e^x \ln(1+y)$	$f_{xx}(0,0) = 0$
$f_{xy}(x,y) = f_{yx} = \dfrac{e^x}{1+y}$	$f_{xy}(0,0) = 1$
$f_{yy}(x,y) = -\dfrac{e^x}{(1+y)^2}$	$f_{yy}(0,0) = -1$

Relying on Taylor's theorem, this leads to

$$\begin{aligned} f(x,y) &= 0+0(x-0)+1(y-0)+\frac{1}{2}\Big(0(x-0)^2+2(x-0)(y-0)+(-1)(y-0)^2\Big)+\cdots \\ &= y+xy-\frac{y^2}{2}+\cdots \end{aligned}.$$

Finally, we have

$$e^x \ln(1+y) = y+xy-\frac{y^2}{2}+\cdots$$

for $y>-1$.

Of course, Taylor series expansions may be extended to functions of n variables.

16.9 Extreme values

Given a function $z=f(x,y)$, we are often interested in points where z takes on the largest or smallest values. For instance, if z represents a cost function, we would likely want to know what (x,y) values minimize the cost. If z represents the ratio of a volume to surface area, we would likely want to know where z is greatest. This leads to the following definition.

Definition 16.21 (Relative and absolute extreme)
Let $z=f(x,y)$ be defined on a set S containing the point $P=(x_0,y_0)$.

1. If $f(x_0,y_0) \geq f(x,y)$ for all (x,y) in S, then f has an **absolute maximum** (*globaal maximum*) at P.

   If $f(x_0,y_0) \leq f(x,y)$ for all (x,y) in S, then f has an **absolute minimum** (*globaal minimum*) at P.

2. If there is an open disk D containing P such that $f(x_0,y_0) \geq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative maximum** (*lokaal maximum*) at P.

   If there is an open disk D containing P such that $f(x_0,y_0) \leq f(x,y)$ for all points (x,y) that are in both D and S, then f has a **relative minimum** (*lokaal minimum*) at P.

3. If f has an absolute maximum or minimum at P, then f has an **absolute extrema** at P.

   If f has a relative maximum or minimum at P, then f has a **relative extrema** at P.

If f has a relative or absolute maximum at $P = (x_0, y_0)$, it means that every curve on the surface of f through P will also have a relative or absolute maximum at P. Recalling what we learned in Section 10.1, the slopes of the tangent lines to these curves at P must be 0 or undefined. Since directional derivatives are computed using f_x and f_y, we are led to the following definition and theorem.

Definition 16.22 (Critical point)
Let $z = f(x, y)$ be continuous on a set S. A **critical point** (*kritisch punt*) $P = (x_0, y_0)$ of f is a point in S such that, at P,

$$f_x(x_0, y_0) = 0 \text{ and } f_y(x_0, y_0) = 0$$

Besides, when $f_x(x_0, y_0)$ and/or $f_y(x_0, y_0)$ is undefined, we call $P = (x_0, y_0)$ a **singular point**, just as we did within the framework of functions of one variable (Definition 10.4).

Theorem 16.15 (Critical and singular points and relative extrema)
Let $z = f(x, y)$ be defined on an open set S containing $P = (x_0, y_0)$. If f has a relative extrema at P, then P is a critical or singular point of f.

Therefore, to find relative extrema, we find the critical and singular points of f and determine which correspond to relative maxima, relative minima, or neither. The following examples demonstrates this.

Example 16.36

Let $f(x, y) = -\sqrt{x^2 + y^2} + 2$. Find the relative extrema of f. The surface of f is graphed in Figure 16.21 along with the point $(0, 0, 2)$.

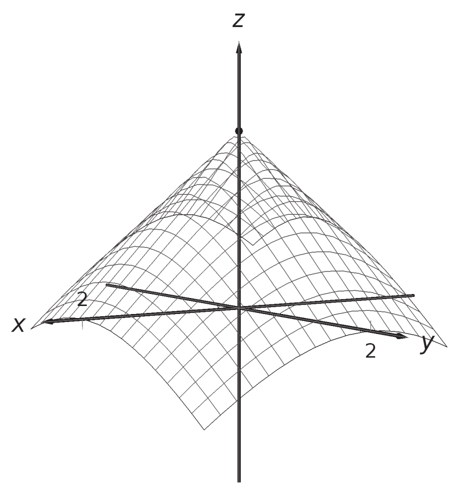


Figure 16.21: The surface in Example 16.36 with its absolute maximum indicated.

Solution

We start by computing the partial derivatives of f:

$$f_x(x, y) = \frac{-x}{\sqrt{x^2 + y^2}} \quad \text{and} \quad f_y(x, y) = \frac{-y}{\sqrt{x^2 + y^2}}.$$

It is clear that $f_x = 0$ when $x = 0$ and $y \neq 0$, and that $f_y = 0$ when $y = 0$ and $x \neq 0$. At $(0, 0)$, both

f_x and f_y are not 0, but rather undefined. The point $(0,0)$ is hence a singular point, though. The graph in Figure 16.21 shows that this point is the absolute maximum of f.

Example 16.37

Let $f(x,y) = x^3 - 3x - y^2 + 4y$. Find the relative extrema of f.

Solution

Once again we start by finding the partial derivatives of f:

$$f_x(x,y) = 3x^2 - 3 \qquad \text{and} \qquad f_y(x,y) = -2y + 4.$$

Each is always defined. Setting each equal to 0 and solving for x and y, we find

$$\begin{aligned} f_x(x,y) = 0 &\quad \Rightarrow \quad x = \pm 1 \\ f_y(x,y) = 0 &\quad \Rightarrow \quad y = 2. \end{aligned}$$

We have two critical points: $(-1,2)$ and $(1,2)$, while there are no singular points. To determine if they correspond to a relative maximum or minimum, we consider the graph of f in Figure 16.22.

The critical point $(-1,2)$ clearly corresponds to a relative maximum. However, the critical point at $(1,2)$ is neither a maximum nor a minimum, displaying a different, interesting characteristic.

If one walks parallel to the y-axis towards this critical point, then this point becomes a relative maximum along this path. But if one walks towards this point parallel to the x-axis, this point becomes a relative minimum along this path. A point that seems to act as both a maximum and a minimum is a saddle point. A formal definition follows.

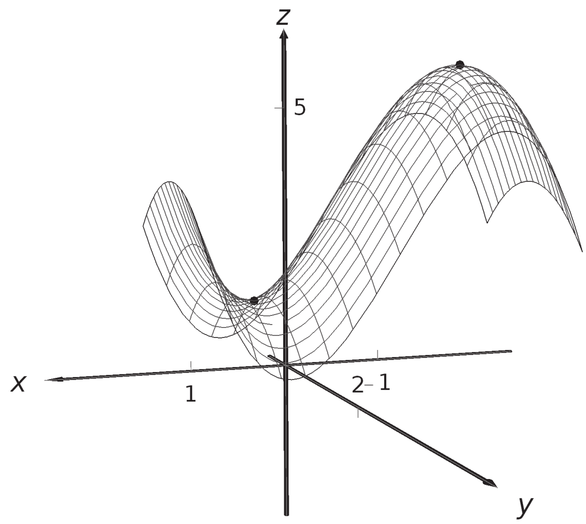


Figure 16.22: The surface in Example 16.37 with both critical points marked.

Definition 16.23 (Saddle point)
Let $P = (x_0, y_0)$ be in the domain of f where $f_x = 0$ and $f_y = 0$ at P. We say P is a **saddle point** (*zadelpunt*) of f if, for every open disk D containing P, there are points (x_1, y_1) and (x_2, y_2) in D such that $f(x_0, y_0) > f(x_1, y_1)$ and $f(x_0, y_0) < f(x_2, y_2)$.

At a saddle point, the instantaneous rate of change in all directions is 0 and there are points nearby with z-values both less than and greater than the z-value of the saddle point.

Before Example 16.37 we mentioned the need for a test to differentiate between relative maxima and minima. We now recognize that our test also needs to account for saddle points. To do so, we consider the second partial derivatives of f. Recall that with single variable functions, such as $y = f(x)$, if $f''(c) > 0$, then if f is concave up at c, and if $f'(c) = 0$, then f has a relative minimum at $x = c$. Note that at a saddle point, it seems the graph is both concave up and concave down, depending on which direction you are considering.

It would be nice if the following were true:

$$\begin{aligned} f_{xx} \text{ and } f_{yy} > 0 \quad &\Rightarrow \quad \text{relative minimum,} \\ f_{xx} \text{ and } f_{yy} < 0 \quad &\Rightarrow \quad \text{relative maximum,} \\ f_{xx} \text{ and } f_{yy} \text{ have opposite signs} \quad &\Rightarrow \quad \text{saddle point.} \end{aligned}$$

However, this is not the case. Functions f exist where f_{xx} and f_{yy} are both positive but a saddle point still exists. In such a case, while the concavity in the x-direction is up (i.e., $f_{xx} > 0$) and the concavity in the y-direction is also up (i.e., $f_{yy} > 0$), the concavity switches somewhere in between the x- and y-directions.

To account for this, consider

$$D = f_{xx}f_{yy} - f_{xy}f_{yx}.$$

Since f_{xy} and f_{yx} are equal when continuous (refer back to Theorem 16.2), we can rewrite this as $D = f_{xx}f_{yy} - f_{xy}^2$. D can be used to test whether the concavity at a point changes depending on direction. If $D > 0$, the concavity does not switch (i.e., at that point, the graph is concave up or down in all directions). If $D < 0$, the concavity does switch. If $D = 0$, our test fails to determine whether concavity switches or not. We state the use of D in the following theorem.

Theorem 16.16 (Second derivative test)
Let R be an open set on which a function $z = f(x, y)$ and all its first and second partial derivatives are defined, let $P = (x_0, y_0)$ be a critical point of f in R, and let

$$D(x_0, y_0) = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0).$$

1. *If $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then f has a relative minimum at P.*
2. *If $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then f has a relative maximum at P.*
3. *If $D < 0$, then f has a saddle point at P.*
4. *If $D = 0$, the test is inconclusive.*

We practice this test with the function in the previous example, where we visually determined we had a relative maximum and a saddle point.

Example 16.38

Let $f(x, y) = x^3 - 3x - y^2 + 4y$ as in Example 16.37. Determine whether the function has a relative minimum, maximum, or saddle point at each critical point.

Solution

We determined previously that the critical points of f are $(-1, 2)$ and $(1, 2)$. To use the second derivative test, we must find the second partial derivatives of f:

$$f_{xx} = 6x; \qquad f_{yy} = -2; \qquad f_{xy} = 0.$$

Thus $D(x,y) = -12x$.

At $(-1,2)$: $D(-1,2) = 12 > 0$, and $f_{xx}(-1,2) = -6$. By the second derivative test, f has a relative maximum at $(-1,2)$.

At $(1,2)$: $D(1,2) = -12 < 0$. The second derivative test states that f has a saddle point at $(1,2)$.

The second derivative test confirmed what we determined visually.

16.10 Exercices

1. Bepaal het domein en het beeld van de onderstaande functies.

   (a) $f(x,y) = \dfrac{\ln(x)}{\sin(y)}$

   (b) $f(x,y) = \dfrac{2+\arcsin(y)}{\ln(2x)}$

   (c) $f(x,y) = \sqrt{1-x^2-y^2}$

   (d) $f(x,y) = \sin(xy)$

   (e) $f(x,y) = |x| - |y|$

   (f) $f(x,y) = \dfrac{1}{x-y}$

   (g) $f(x,y) = \ln^{-1}(x^2+y^2-3)$

   (h) $f(x,y) = \pi - \arcsin(x^2+2y^2)$

2. Schets enkele niveaukrommes van de onderstaande functies.

   (a) $f(x,y) = \dfrac{x^2}{y}$

   (b) $f(x,y) = \dfrac{y}{x^2+y^2}$

   (c) $f(x,y) = \sqrt{\dfrac{1}{y} - x^2}$

3. Beschrijf telkens de grafiek van de functie $f(x,y)$ waarvan de niveaukrommes $f(x,y) = C$ worden voorgesteld in Figuur 16.23 (a) en (b). De grootste cirkel in Figuur 16.23 (a) heeft als straal $r = 5$.

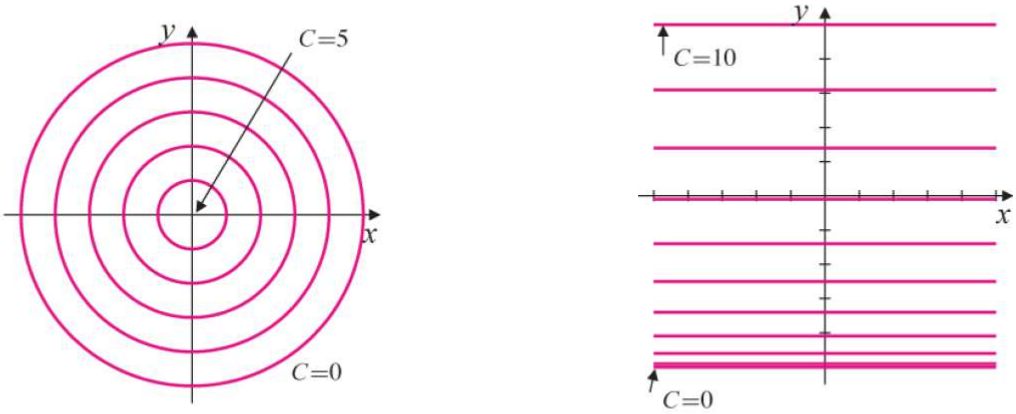


Figure 16.23: Niveaukrommes $f(x,y) = C$ van de functies $f(x,y)$ uit oefening 3.

4. Bepaal $f_x(0,0)$ en $f_y(0,0)$ van de onderstaande functie, indien ze bestaan, met behulp van Definitie 16.7.

$$f(x,y) = \begin{cases} (x^3+y)\sin\left(\dfrac{1}{x^2+y^2}\right), & \text{als } (x,y) \neq (0,0), \\ 0, & \text{als } (x,y) = (0,0). \end{cases}$$

Bereken $f_x(x,y)$. Is deze continu in $(0,0)$?

5. Bereken de partiële afgeleiden van de eerste en tweede orde van de onderstaande functies en dit telkens in het aangegeven punt. Bepaal eveneens de totale differentialen van de eerste en tweede orde in het vermelde punt.

   (a) $f(x,y) = x^2+2xy+y^2-2x+3y-7$ $\quad(1,2)$
   (b) $f(x,y) = x^2y^5+xy^2+x^3y$ $\quad(3,1)$
   (c) $f(x,y) = \dfrac{xy}{x^2+y^2}$ $\quad(1,1)$
   (d) $f(x,y) = x^y$ $\quad(1,1)$
   (e) $f(x,y) = \ln(2x-3y)$ $\quad(2,1)$
   (f) $f(x,y) = \dfrac{e^y}{x}$ $\quad(1,1)$
   (g) $f(x,y) = e^{\frac{y}{x}}$ $\quad(1,1)$
   (h) $f(x,y) = \cos(3x+2y)$ $\quad(0,\pi)$
   (i) $f(x,y) = \arctan(x+y)$ $\quad(1,0)$
   (j) $f(x,y) = x\sinh(y)+y\cosh(x)$ $\quad(0,0)$
   (k) $f(x,y) = (2x+y)^{x+3y}$ $\quad(0,1)$

6. Bereken de partiële afgeleiden van de eerste en tweede orde van de onderstaande functies en dit telkens in het aangegeven punt.

   (a) $f(x,y,z) = \arctan(x+y+z)$ $\quad(1,0,0)$
   (b) $f(x,y,z) = x^2+3y^2+6z^2-2xy+6xz+7yz+4x-3y+7$ $\quad(0,0,0)$
   (c) $f(x,y,z) = \sqrt{xy+z^2}$ $\quad(1,1,1)$
   (d) $f(x,y,z) = xe^{xy+z}$ $\quad(1,0,0)$
   (e) $f(x,y,z) = e^{x+y^2+z^3}$ $\quad(0,0,0)$
   (f) $f(x,y,z) = x\sin(y)+y\ln(z)$ $\quad(1,\pi,1)$
   (g) $f(x,y,z) = (xy)^z+z^{xy}$ $\quad(1,1,1)$

7. We beschouwen $z=\ln\sqrt{x^2+y^2}$. Toon aan dat
$$x\,\frac{\partial z}{\partial x}+y\,\frac{\partial z}{\partial y}=1.$$

8. We beschouwen
$$z=e^{\frac{x}{y}}\sin\left(\frac{x}{y}\right)+e^{\frac{y}{x}}\cos\left(\frac{y}{x}\right).$$
   Toon aan dat
$$x\,\frac{\partial z}{\partial x}+y\,\frac{\partial z}{\partial y}=0.$$

9. Bewijs dat $z=f(x)g(y)$ voldoet aan
$$z\,\frac{\partial^2 z}{\partial x\partial y}=\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}.$$

10. We beschouwen $z=A\sin(\alpha\lambda y+\varphi)\sin(\lambda x)$. Toon aan dat
$$\frac{\partial^2 z}{\partial y^2}=\alpha^2\,\frac{\partial^2 z}{\partial x^2}.$$

11. Toon aan dat de functie $u(x,y,t) = t^{-1}\, e^{-(x^2+y^2)/(4t)}$ voldoet aan de twee-dimensionale warmte-vergelijking
$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

12. We beschouwen $u = x^2y + y^2z + z^2x$. Toon aan dat
$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} = (x+y+z)^2.$$

13. De zijden van een rechthoekige doos zijn tot op 1% van hun lengte nauwkeurig gemeten. Hoeveel is het geschatte maximale foutenpercentage van
    (a) het volume van de doos,
    (b) de oppervlakte van een zijkant van de doos,
    (c) de lengte van een diagonaal van de doos?

14. We beschouwen $z = xy$ met $x = \dfrac{1}{t}$ en $y = t^2$. Bepaal $\frac{dz}{dt}$.

15. Bepaal u_t als $u = \sqrt{x^2+y^2}$ met $x = e^{st}$ en $y = 1 + s^2\cos(t)$.

16. We beschouwen $z = x^2 + 2xy + 4y^2$ met $y = e^{ax}$. Bepaal $\frac{dz}{dx}$.

17. We beschouwen $z = f(u,v)$ met $u = u(x,y)$ en $v = v(x,y)$. Bepaal z_x en z_y.
    (a) $z = \ln\left(u^2+v^2\right)$ met $\begin{cases} u = x+2y+1 \\ v = 3x-y-1 \end{cases}$
    (b) $z = \cosh\left(u^2-v^2\right)$ met $\begin{cases} u = 2x-3y \\ v = 3x-4y \end{cases}$
    (c) $z = \dfrac{v}{u}$ met $\begin{cases} u = \sin\left(x^2-y^2\right) \\ v = e^{xy} \end{cases}$
    (d) $z = ve^{uv}$ met $\begin{cases} u = x^2y \\ v = xy^2 \end{cases}$
    (e) $z = u^v$ met $\begin{cases} u = x^2+y^2 \\ v = xy \end{cases}$

18. We beschouwen $z = f(x,y)$ met $x = 2s+3t$ en $y = 3s-2t$. Bepaal
$$\frac{\partial^2 z}{\partial s^2}, \quad \frac{\partial^2 z}{\partial s \partial t} \quad \text{en} \quad \frac{\partial^2 z}{\partial t^2}.$$

19. Beschouw $x = e^s\cos(t)$, $y = e^s\sin(t)$ en $z = u(x,y) = v(s,t)$. Toon aan dat
$$\frac{\partial^2 z}{\partial s^2} + \frac{\partial^2 z}{\partial t^2} = (x^2+y^2)\left(\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}\right).$$

20. Beschouw $u(x,y) = r^2\ln(r)$, met $r^2 = x^2+y^2$. Ga na dat u een biharmonische functie is door aan te tonen dat
$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = 0.$$

21. Beschouw de functie $w = f(x,y,z)$ met $x = g(s)$, $y = h(s,t)$ en $z = k(t)$. Bepaal een uitdrukking voor $\frac{\partial w}{\partial t}$.

22. Beschouw de functie $z = g(x,y)$ met $y = f(x)$ en $x = h(u,v)$. Bepaal een uitdrukking voor $\frac{\partial z}{\partial u}$.

23. Beschouw de functie $w = f(x,y)$ met $x = g(r,s)$, $y = h(r,t)$, $r = k(s,t)$ en $s = m(t)$. Bepaal een uitdrukking voor $\frac{dw}{dt}$.

24. Bereken telkens de gevraagde afgeleide voor de gegeven vergelijking. Welke voorwaarde(n) voor de veranderlijken zal het bestaan van de aangegeven functie (oplossing van de desbetreffende afgeleide) verzekeren?

    (a) $xy^3 + x^4y = 2$ definieert x als een functie van y. Bepaal dan $\frac{dx}{dy}$.

    (b) $z^2 + xy^3 = \dfrac{xz}{y}$ definieert z als een functie van x en y. Bepaal dan $\frac{\partial z}{\partial y}$.

    (c) $e^{yz} - x^2 z \ln(y) = \pi$ definieert y als een functie van x en z. Bepaal dan $\frac{\partial y}{\partial z}$.

25. Door de impliciete vergelijking $x^2 + y^2 = r^2$ worden twee functies $y = f_1(x)$ en $y = f_2(x)$ bepaald.

    (a) Bewijs dat de vergelijking van de raaklijn aan de grafiek in het punt $P = (x_1, y_1)$ van de grafiek $xx_1 + yy_1 = r^2$ is.

    (b) Bepaald de richtingscoëfficiënt van die raaklijn en toon aan dat die raaklijn loodrecht staat op het lijnstuk $[OP]$.

26. Heeft de vergelijking $(x^2 + y^2 + 2z^2)^{1/2} = \cos(z)$ een unieke oplossing voor y in functie van x en z in de buurt van $(0,1,0)$? Bestaat er een unieke oplossing voor z in functie van x en y?

27. Veronderstel dat $F(x,y) = 0$ van de klasse C^1 is en dat $F(0,0) = 0$. Welke voorwaarden voor F zorgen ervoor dat $F(F(x,y),y) = 0$ kan opgelost worden naar y als een C^1 functie van x in de buurt van $(0,0)$?

28. We beschouwen $f(x,y) = \ln(x^2 + y^2)$ en $P = (1,-2)$. Bepaal

    (a) de gradiënt van de gegeven functie in het gegeven punt,

    (b) een vergelijking van het raakvlak aan de grafiek van de gegeven functie in het gegeven punt,

    (c) een vergelijking van de raaklijn in het gegeven punt aan de niveaukromme van de gegeven functie door het gegeven punt.

29. Bepaal een vergelijking van het raakvlak aan het niveauoppervlak van de functie $f(x,y,z) = \cos(x + 2y + 3z)$ in het punt $(\pi/2, \pi, \pi)$.

30. Bepaal de mate van verandering van de functie $f(x,y) = x^2y$ in het punt $(-1,-1)$ in de richting van $\vec{\boldsymbol{v}} = (1,2)$.

31. We beschouwen $f(x,y) = x + y^2 - 3xy + 5y - 1$. In welke richting in het punt $(1,1)$ verandert de functiewaarde het meest?

32. Bepaal de directionele afgeleide $D_{\vec{\boldsymbol{u}}}f(P)$ als $f(x,y,z) = xy^2z^3$, $P = (1,1,1)$ en $\vec{\boldsymbol{u}}$ loodrecht staat op het oppervlak $x^4 + 2y^4 + 2z^4 = 5$ in P.

33. We beschouwen

$$G(x,y,z) = z \ln\left(\frac{xz+1}{y+2}\right)$$

    en verder $\vec{\boldsymbol{a}} = (1,2,1)$ en $\vec{\boldsymbol{b}} = (5,4,-3)$.
    Bepaal de directionele afgeleide $D_{\vec{\boldsymbol{u}}}G(\vec{\boldsymbol{a}})$ waarbij $\vec{\boldsymbol{u}}$ dezelfde richting heeft als de richting van $\vec{\boldsymbol{a}}$ naar $\vec{\boldsymbol{b}}$.

34. De temperatuur $T(x,y)$ in punten in het (x,y)-vlak wordt gegeven door $T(x,y) = x^2 - 2y^2$.

    (a) Teken een contourplot voor T bestaande uit isothermen (krommes van constante temperatuur).

    (b) In welke richting moet een mier in $(2,-1)$ bewegen om zo snel mogelijk af te koelen?

    (c) Langs welke kromme door het punt $(2,-1)$ moet de mier bewegen om maximale afkoeling te blijven ervaren?

35. Een renpaard leeft in een vallei die gemodelleerd wordt door $f(x,y) = 3x^2 + y^2$. Het paard staat op een renbaan die de verzameling is van punten waarvoor geldt dat $x^2 + y^2 = 1$. Het renpaard heeft als wens om te ontsnappen uit de vallei en vanaf de renbaan het steilst mogelijke pad te nemen naar boven. In welk punt moet het paard ontsnappen en in welke richting moet het vanuit dat punt naar boven lopen?

36. Bepaal de vergelijkingen van het raakvlak en de normaal aan de grafiek van de gegeven functie in het gegeven punt.

    (a) $f(x,y) = \dfrac{2xy}{x^2+y^2}$ in $(0,2)$

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

    (c) $2x^2 + 4yz - 5z^2 = -10$ in $(3,-1,2)$

    (d) $f(x,y) = 2\sin(x)\cos(y)$ in $(\pi/4, \pi/4, 1)$

37. Bepaal de afstand van het punt $(1,1,0)$ tot de cirkelvormige paraboloïde $z = x^2 + y^2$.

38. Bepaal van de onderstaande functies een Taylor-reeksontwikkeling van de tweede orde rond het opgegeven punt.

    (a) $f(x,y) = xe^{xy+y}$ $(1,1)$

    (b) $f(x,y) = x\ln(y)$ $(0,1)$

    (c) $f(x,y) = xy + \ln(xy)$ $(1,1)$

    (d) $f(x,y) = x\sin(y)$ $(1,0)$

    (e) $f(x,y) = xy\cos(x+y)$ $\left(0, \dfrac{\pi}{2}\right)$

    (f) $f(x,y) = \arctan\left(\dfrac{x}{y}\right)$ $(0,1)$

    (g) $f(x,y) = \sin(xe^y)$ $(1,0)$

    (h) $f(x,y) = \dfrac{\sin(x)}{y}$ $\left(\dfrac{\pi}{2}, 1\right)$

39. Geef een benadering voor

    (a) $\arctan\left(\dfrac{1.02}{0.95}\right)$,

    (b) $\sqrt{3.99 \times 4.02}$.

40. Bepaal en classificeer de extrema van de onderstaande functies.

    (a) $z = x^2 + y^2 - 3xy$

    (b) $z = xy$

    (c) $z = y\sqrt{x} - xy + y^2$

    (d) $z = (2x^2 - y)(2 - y)$

    (e) $z = x^4 + y^4 - 2x^2 + 4xy - 2y^2$

    (f) $z = x^3 - 3xy^2$

    (g) $z = (x-y)^4 + (y-1)^4$

    (h) $z = x + y\sin(x)$

(i) $z = (x+y)e^{-xy}$

(j) $z = x^2 + y^2$

(k) $z = \cos(x+y)$