COMPLEX
SEQUENCES & SERIES

Course notes for Math 200

Scott Corry

Lawrence University
Appleton, WI

VERSION: June 22, 2020

. . . you are surprised at my working simultaneously in literature and in mathematics. Many people who have never had occasion to learn what mathematics is confuse it with arithmetic and consider it a dry and arid science. In actual fact it is the science which demands the utmost imagination. One of the foremost mathematicians of our century says very justly that it is impossible to be a mathematician without also being a poet in spirit. It goes without saying that to understand the truth of this statement one must repudiate the old prejudice by which poets are supposed to fabricate what does not exist, and that imagination is the same as "making things up." It seems to me that the poet must see what others do not see, and see more deeply than other people. And the mathematician must do the same.

Sofia Kovalevskaya, 1890

It has been written that the shortest and best way between two truths of the real domain often passes through the imaginary one.

Jacques Hadamard, 1945

CONTENTS

ABOUT THE NOTES

These notes were written during my sabbatical year 2018-19 as part of a restructuring of the first-year calculus curriculum at Lawrence University. The plan for MATH 200 took shape through conversations with my Lawrence colleagues Alan Parks, Julie Rana, and Liz Sattler, and their comments and suggestions have had a strong influence. I particularly want to acknowledge Liz Sattler, who read multiple drafts of each chapter and provided substantive and detailed feedback, ranging from high-level organizational and expository suggestions, through fine-grained comments on argument and notation, down to the meticulous correction of typos and figure placement. She also contributed a large portion of the end-of-chapter problems. Her hard work and excellent pedagogical sense have improved the quality of these notes immensely. We hope the students enjoy them.

Scott Corry
Valencia, Spain
March, 2019

CHAPTER 1

NUMBERS

1.1. Complex Numbers

Real numbers provide the context of the calculus you have learned so far, both one- and multi-variable. This course will develop some new topics in the context of the *complex* numbers, commonly used in pure and applied mathematics, physics, and engineering. We begin by giving a quick introduction to the algebra of these numbers, and then spend the rest of this section describing their geometry.

Start by introducing a new number i with the property that

$$i^2 = -1.$$

This is the famous square root of -1 that you may have heard about. Consider all expressions of the form $a+bi$, where a and b are real numbers. Now do algebra with these expressions following the usual rules of associativity, distributivity, and commutativity, always remembering that $i^2 = -1$:

$$\begin{aligned} \textbf{Addition:} \qquad (a+bi)+(c+di) &= (a+c)+(b+d)i \\ \textbf{Multiplication:} \qquad (a+bi)(c+di) &= ac+adi+bci+bdi^2 \\ &= (ac-bd)+(ad+bc)i \end{aligned}$$

DEFINITION 1.1. Expressions $z = a+bi$ are called *complex numbers*. The real number a is the *real part* of z and denoted $\mathrm{Re}(z)$. Similarly, the real number b is the *imaginary part* of z and denoted $\mathrm{Im}(z)$. We use the symbol $\mathbb{C}$ to indicate the set of all complex numbers.

EXAMPLE 1.2. Suppose that $z = 3+2i$ and $w = 1-5i$. Then

$$\begin{aligned} z+w &= (3+2i)+(1-5i) \\ &= (3+1)+(2-5)i \\ &= 4-3i \end{aligned}$$

$$\begin{aligned} zw &= (3+2i)(1-5i) \\ &= 3-15i+2i-10i^2 \\ &= (3+10)+(-15+2)i \\ &= 13-13i. \end{aligned}$$

EXAMPLE 1.3. For another example, suppose that $z = 1-\sqrt{2}i$ and $w = \frac{1}{3}+\frac{\sqrt{2}}{3}i$. Then

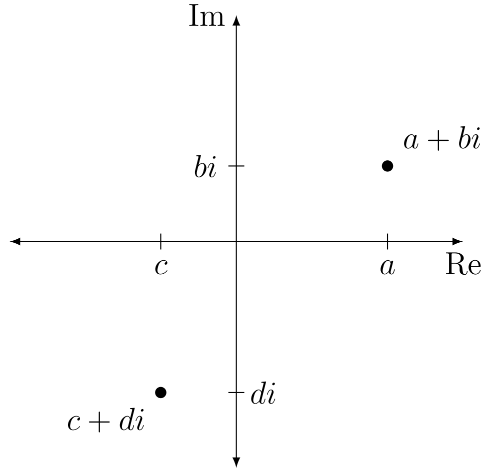
$$\begin{aligned} z+w &= (1-\sqrt{2}i)+\left(\frac{1}{3}+\frac{\sqrt{2}}{3}i\right) \\ &= \left(1+\frac{1}{3}\right)+\left(-\sqrt{2}+\frac{\sqrt{2}}{3}\right)i \\ &= \frac{4}{3}-\frac{2\sqrt{2}}{3}i \end{aligned}$$

and

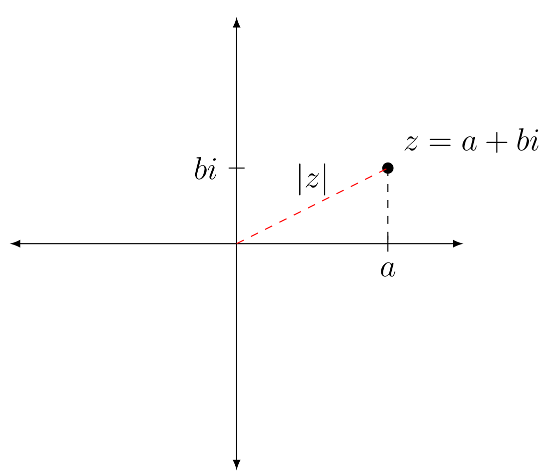
$$\begin{aligned} zw &= (1-\sqrt{2}i)\left(\frac{1}{3}+\frac{\sqrt{2}}{3}i\right) \\ &= \frac{1}{3}+\frac{\sqrt{2}}{3}i-\frac{\sqrt{2}}{3}i-\frac{2}{3}i^2 \\ &= \left(\frac{1}{3}+\frac{2}{3}\right)+\left(\frac{\sqrt{2}}{3}-\frac{\sqrt{2}}{3}\right)i \\ &= 1+0i \\ &= 1. \end{aligned}$$

REMARK 1.4. The pejorative term “imaginary” is a holdover from a time when people were skeptical about the existence of these numbers, first introduced by Italian mathematicians in the 16th century in their efforts to solve polynomial equations. If you also feel a bit uneasy at this point, don't worry: one of the main goals of this course is for you to become as comfortable with complex numbers as you are with real numbers. By the end, we hope you will agree that complex numbers are no more imaginary than any other numbers.

We visualize the complex numbers $\mathbb{C}$ as a plane, with horizontal axis corresponding to the real part, and vertical axis to the imaginary part:



Definition 1.5. If $z = a + bi$ is a complex number, then its *magnitude* $|z| = \sqrt{a^2 + b^2}$ is the distance from z to the origin. The word *modulus* is also often used for the magnitude of a complex number.



The proposition below records the fact that the magnitude of a complex number z is greater than or equal to the magnitude of $\operatorname{Re}(z)$ and the magnitude of $\operatorname{Im}(z)$.

PROPOSITION 1.6. *Suppose that $z = a + bi$ is a complex number. Then*

$$|z| \geq |a| \quad \textit{and} \quad |z| \geq |b|.$$

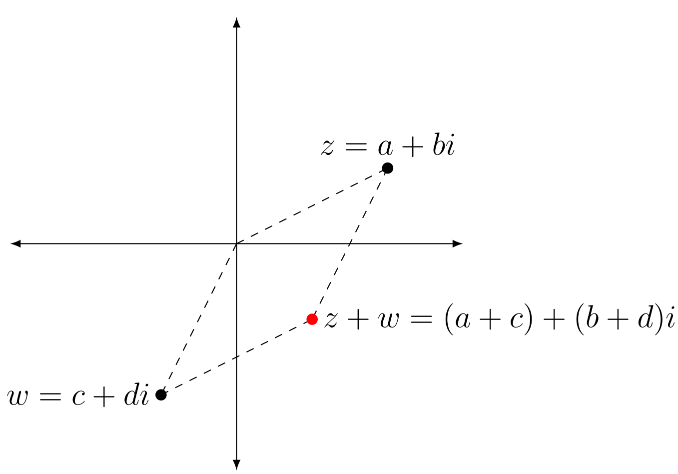
PROOF. Geometrically, this is just the statement that the hypotenuse of a right triangle is longer than each leg (see picture above). Here is the corresponding algebraic argument for the real part a:

$$|z|^2 = a^2 + b^2 \geq a^2 = |a|^2.$$

Taking the square root of both sides yields $|z| \geq |a|$ as claimed. A similar argument shows that $|z| \geq |b|$. □

What do the operations of addition and multiplication look like geometrically? For addition, we have the parallelogram law, familiar from vector addition in $\mathbb{R}^2$.

Parallelogram Law for Complex Addition: the sum of $z = a + bi$ and $w = c + di$ is the vertex across from the origin in the parallelogram shown below:



The following result tells us how the magnitude relates to complex addition. While simple, it is of fundamental importance, and will be used several times throughout the course:

PROPOSITION 1.7 (The Triangle Inequality). *Suppose that z and w are complex numbers. Then*

$$|z + w| \leq |z| + |w|.$$

PROOF. We will content ourselves with a geometric justification; see Problem 1.7 for an algebraic argument. Consider the triangle shown in Figure 1.1, formed from half of the parallelogram determined by z and w. The picture shows that the stated inequality is basically just the assertion that "the shortest path between two points in the plane is the straight line." □

EXERCISE 1.1. When does equality occur in the statement of the triangle inequality?

In order to understand the geometric meaning of complex multiplication, we need to think about the plane in polar coordinates.

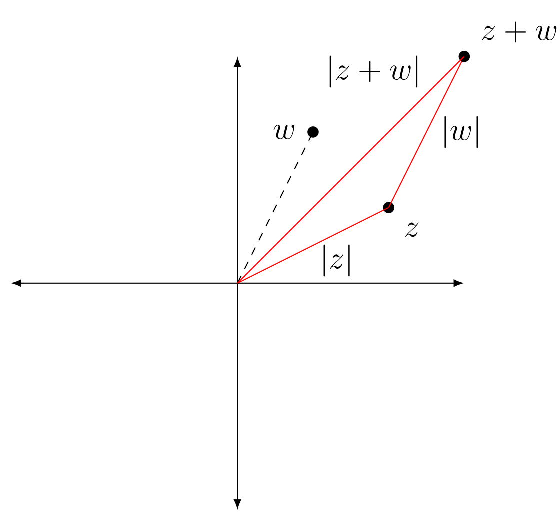


FIGURE 1.1. The triangle inequality

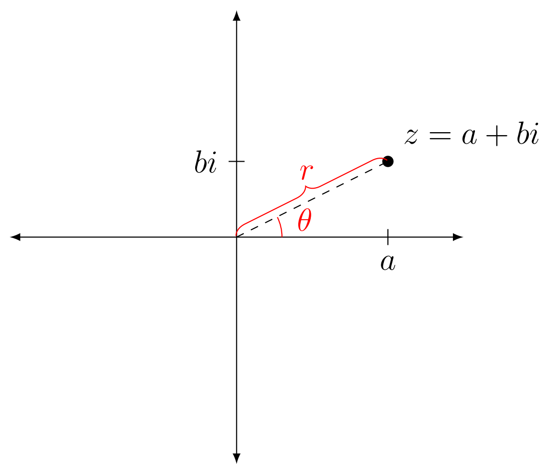
DEFINITION 1.8. If z is a complex number, then the *polar coordinates* of z are given by the pair of real numbers (r, θ) where

$$r = \text{distance from the origin} = |z|$$

and

$$\theta = \text{angle from the positive real axis.}$$

The angle θ is called the *argument* of z and denoted $\arg(z)$. Note that the argument θ may be replaced by $\theta' = \theta + 2\pi n$ for any integer n without changing the number z. The term *phase* is also often used for the argument of a complex number, especially by physicists.



Trigonometry and the Pythagorean Theorem allow us to determine the real and imaginary parts of z if we know its polar coordinates, and vice-versa. Referring to the picture above:

$$a = r\cos(\theta) \quad , \quad b = r\sin(\theta)$$

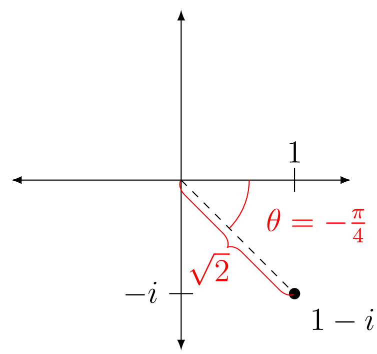
and

$$r = \sqrt{a^2 + b^2} \quad , \quad \theta = \begin{cases} \arctan(b/a) & \text{if } a > 0 \\ \arctan(b/a) + \pi & \text{if } a < 0 \\ \pi/2 & \text{if } a = 0 \text{ and } b > 0 \\ -\pi/2 & \text{if } a = 0 \text{ and } b < 0. \end{cases}$$

Here, $\arctan\colon \mathbb{R} \to (-\pi/2, \pi/2)$ is the inverse of the tangent function.

DEFINITION 1.9. Let z be a complex number. Writing $z = a + bi$ expresses z in its *cartesian form*. The *polar form* of z is

$$z = r\cos(\theta) + ir\sin(\theta) = |z|(\cos(\theta) + i\sin(\theta)).$$



EXAMPLE 1.10. Let's find the polar form of the complex number $z = 1 - i$, shown on the picture above. The magnitude is given by

$$|z| = \sqrt{1^2 + (-1)^2} = \sqrt{2}.$$

Since the real part of z is positive, the argument is

$$\theta = \arctan(-1/1) = \arctan(-1) = -\frac{\pi}{4}.$$

So the polar form of z is

$$z = \sqrt{2}(\cos(-\pi/4) + i\sin(-\pi/4)).$$

Note that we can equally well use the argument $-\pi/4 + 2\pi = 7\pi/4$ and write

$$z = \sqrt{2}(\cos(7\pi/4) + i\sin(7\pi/4)).$$

The next exercise reveals the geometric meaning of complex multiplication.

EXERCISE 1.2. Recall the algebraic formula for the product of two complex numbers $a + bi$ and $c + di$:

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

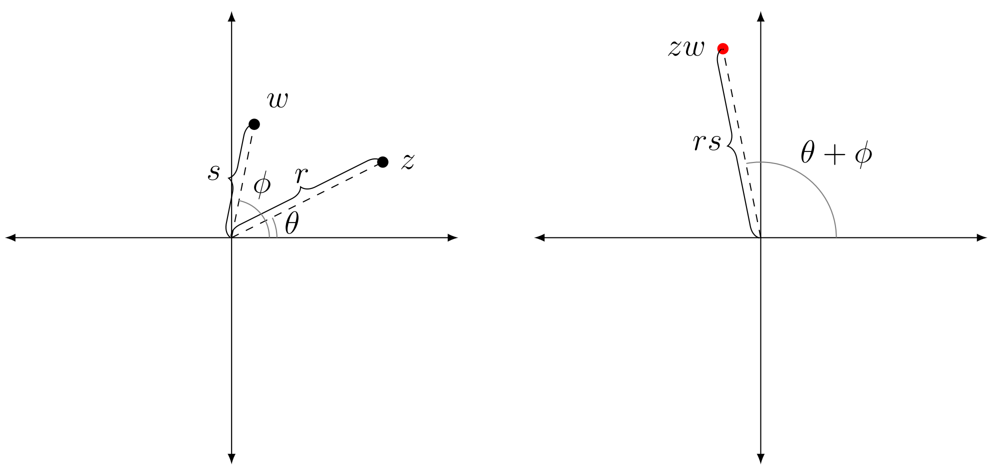
(a) Express two generic complex numbers z and w in polar form, and then use the formula above to compute the product zw.

(b) Check that the $\arg(zw) = \arg(z) + \arg(w)$. You will need to make use of the trigonometric identities

$$\begin{aligned}\cos(\theta+\phi) &= \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi)\\ \sin(\theta+\phi) &= \cos(\theta)\sin(\phi) + \sin(\theta)\cos(\phi).\end{aligned}$$

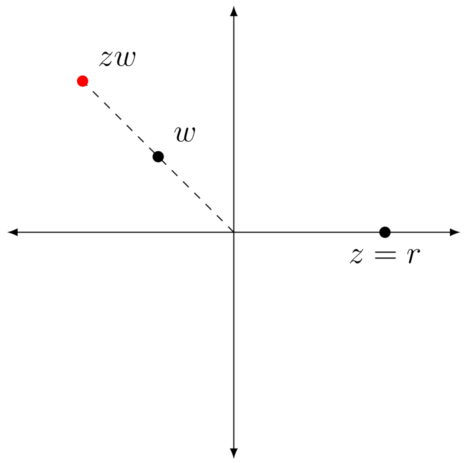
(c) Finally, check that $|zw| = |z||w|$.

Rotation-scale Law for Complex Multiplication: Suppose that z has polar coordinates (r, θ) and w has polar coordinates (s, ϕ). Then the product of z and w is the number zw with polar coordinates $(rs, \theta + \phi)$. In words: when multiplying complex numbers, we multiply the magnitudes and add the arguments. See the picture below.

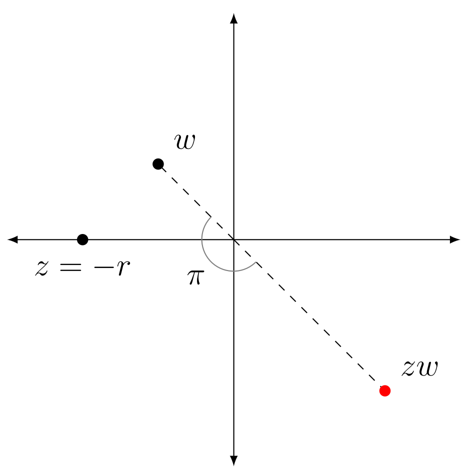


In general, if z has polar coordinates (r, θ), then the effect on w of multiplication by z is to rotate w by θ and scale its magnitude by r. Let's think about the effect of multiplying a fixed complex number w by various types of complex numbers z:

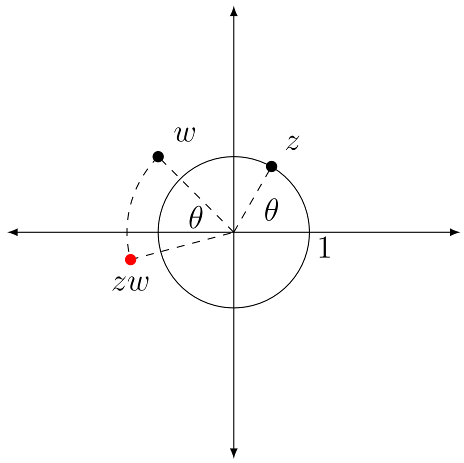
- If $\arg(z) = 0$, then z is on the positive real axis at a distance r from the origin, and multiplication by z simply scales the magnitude of w by the positive real number r:



- If $\arg(z) = \pi$, then z is on the negative real axis at a distance r from the origin, and multiplication by z rotates w by π and scales its magnitude by r:



- If z has magnitude $r = 1$, then z is on the unit circle, and multiplication by z just rotates w by the angle $\theta = \arg(z)$:



REMARK 1.11. Note that if both z and w lie on the real axis, then complex multiplication is just the usual multiplication of real numbers, and for this reason we think of the real axis as a copy of the real numbers, sitting inside the complex plane.

In Example 1.3, we saw that

$$(1-\sqrt{2}i)\left(\frac{1}{3}+\frac{\sqrt{2}}{3}i\right)=1.$$

This means that these two complex numbers are *inverses*, and so we write

$$(1-\sqrt{2}i)^{-1}=\frac{1}{1-\sqrt{2}i}=\frac{1}{3}+\frac{\sqrt{2}}{3}i.$$

In fact, it is not hard to see that every nonzero complex number z has an inverse. Indeed, suppose that z has polar coordinates (r,θ). Since $z\neq 0$, we know that $r>0$, so that $r^{-1}=1/r$ exists as a real number. We are looking for another complex number w so that $zw=1$. But the number 1 has magnitude 1 and argument 0, so we can achieve our goal if we choose w to have polar coordinates $(r^{-1},-\theta)$, because then the polar coordinates of zw will be $(rr^{-1},\theta+(-\theta))=(1,0)$.

So we now know how to find inverses using polar coordinates (see the left side of Figure 1.2):

if $z\neq 0$ has polar coordinates (r,θ), then
z^{-1} has polar coordinates $(r^{-1},-\theta)$.

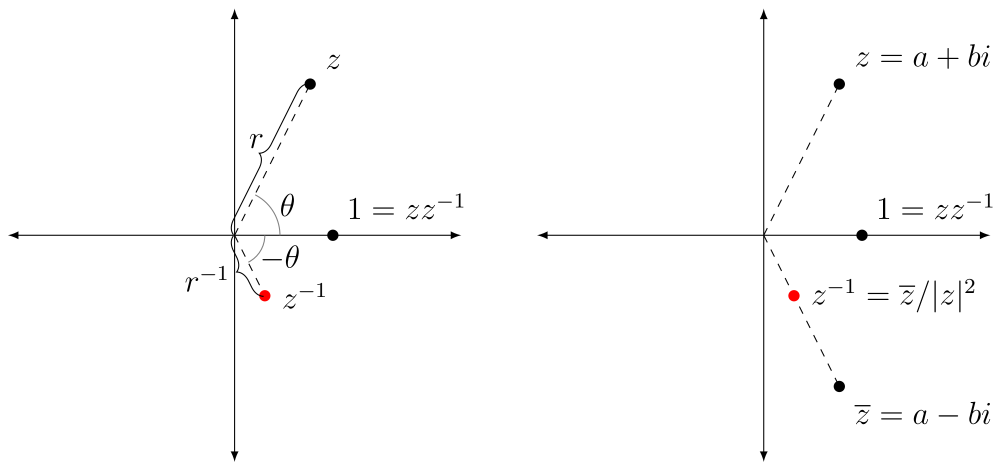


FIGURE 1.2. Finding the inverse using polar coordinates (left) and complex conjugation (right)

EXERCISE 1.3. Can you convince yourself that every nonzero complex number z has exactly one inverse, i.e., that inverses are unique?

How do we find inverses using cartesian coordinates? To answer this, it will be helpful to introduce an important operation on complex numbers.

DEFINITION 1.12. The number $\overline{z} = a - bi$ is called the *complex conjugate* of $z = a + bi$. Geometrically, complex conjugation is simply reflection across the horizontal real axis (see right side of Figure 1.2).

EXERCISE 1.4. Show that $z\overline{z} = |z|^2$, so that $z^{-1} = \overline{z}/|z|^2$.

We can interpret the previous exercise geometrically as follows, using the right side of Figure 1.2:

(1) Starting with $z = a + bi$, reflect across the real axis to obtain the complex conjugate $\overline{z} = a - bi$; this complex number has the same magnitude as z, but the opposite argument $-\theta$.
(2) Now scale $\overline{z}$ to obtain z^{-1} with magnitude $|z|^{-1}$. Since $\overline{z}$ has magnitude $|z|$, we must divide by $|z|^2 = a^2 + b^2$.

Putting these two steps together yields an explicit formula for the inverse in cartesian coordinates:

$$z^{-1} = \frac{\overline{z}}{|z|^2} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2} i.$$

EXAMPLE 1.13. To find the inverse of $z = 3 + \sqrt{2}i$, we divide the complex conjugate by the magnitude squared:

$$z^{-1} = \frac{\overline{z}}{|z|^2} = \frac{3 - \sqrt{2}i}{3^2 + 2} = \frac{3}{11} - \frac{\sqrt{2}}{11} i.$$

EXAMPLE 1.14. Let's write the complex fraction $(1+2i)/(1-i)$ in the cartesian form $a+bi$. First of all, note that division by the complex number $1-i$ means multiplication by its inverse $(1-i)^{-1}$. So we begin by finding this inverse. We have

$$(1-i)^{-1} = \frac{\overline{1-i}}{|1-i|^2} = \frac{1+i}{1^2 + (-1)^2} = \frac{1}{2} + \frac{1}{2} i.$$

It follows that

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

Your initial introduction to calculus took place entirely in the context of the real numbers $\mathbb{R}$, and this course will take place mainly in the context of the complex numbers $\mathbb{C}$. Geometrically, this shift represents a substantial enlargement, since we view the complex numbers as a two-dimensional plane, with a copy of the 1-dimensional real numbers $\mathbb{R}$ inside as the horizontal axis. But at the most mundane symbolic level, not much has really changed: we can add, subtract, multiply, and divide complex numbers, and all the usual rules of algebra apply (associativity, distributivity, and commutativity).

If all this feels a bit strange, don't worry—you will get used to the complex numbers soon enough. But you might be feeling that things are too good to be true: on page 1 we simply assumed the existence of a number i with the property that $i^2 = -1$, forged ahead with the rules of ordinary algebra, and everything turned out fine. What is to stop us from inventing anything we want? Well, nothing prevents us from investigating the consequences of various hypotheses, and one of the most powerful and charming aspects of mathematics is the opportunities for creativity so afforded. This creative aspect of mathematics is often hidden from beginning students, who understandably tend to see mathematics as having fixed and inflexible rules. The truth is that the creativity proper to mathematics is of a highly constrained type: while we are free to make definitions and hypotheses, most will lead nowhere of interest or will yield inconsistencies. The magic of the complex numbers is that they enlarge the real numbers in just the right way to provide a rich computational system, with operations that have a strong geometric interpretation. Moreover, complex numbers are extremely useful, having applications not only within pure mathematics but also in many applied areas such as quantum mechanics and electrical engineering.

As a cautionary and inspiring tale, consider the 19th century Irish mathematician W.R. Hamilton. Motivated by the utility of the complex number system for working in two dimensions, he wanted to similarly find a way of multiplying triples of real numbers so that it would be possible to divide by nonzero triples. (Note that the cross product of vectors doesn't allow for division, since $\mathbf{v} \times \mathbf{w} = \mathbf{0}$ whenever $\mathbf{v}$ and $\mathbf{w}$ are parallel vectors in $\mathbb{R}^3$). He began in the same way we did above, by considering expressions of the form $a + bi + cj$ with a, b, c real and i, j independent elements satisfying $i^2 = j^2 = -1$. But it turns out that "forging ahead" with the usual algebra simply doesn't work, as there is no way to define the product ij without running into inconsistencies. After a decade of work, Hamilton had a flash of insight leading to a major breakthrough: although it is not possible to define a multiplication and division on $\mathbb{R}^3$, it *is* possible to do it for $\mathbb{R}^4$. The resulting 4-dimensional number system is called the *quaternions*, and it satisfies all the usual algebraic properties except for one: multiplication is not commutative. Far from being an exotic curiosity, quaternions have many applications, for instance to the description of rotations in computer graphics. It turns out that $\mathbb{R}^8$ can also be made into something like a number system with division (called the *octonions*), except that the multiplication is neither commutative nor associative! But that is the end of the line: in a precise sense, the only dimensions for which a reasonable notion of multiplication exists (so that nonzero elements have inverses) are 1,2,4, and 8. This startling fact has deep connections to other areas of mathematics and physics—we encourage you to look into it for yourself.

Key points from Section 1.1:

- Cartesian and polar forms of a complex number (Definition 1.9)
- Ability to transform between cartesian and polar forms (page 6 and Example 1.10)
- Visualization of complex addition as parallelogram law (page 4)
- Visualization of complex multiplication as rotation-scaling (page 8)
- Inverse of a complex number (pages 11–12 and Example 1.13)

1.2. Roots of Polynomials

We have introduced the complex numbers and started to get comfortable with their algebra and geometry. We now want to talk about a sense in which the complex numbers are "better" than the real numbers, as a way of justifying their use.

Complex numbers first arose in the effort to find roots of polynomials (recall that a *root* of a polynomial $p(z)$ is a number a such that $p(a) = 0$). You are probably aware that not all polynomials with real coefficients have real roots. For instance, consider the quadratic polynomial $z^2 + 1$. For any real number a, substituting $z = a$ yields the number a^2+1, which is never zero (in fact, it is always at least 1, since $a^2 \geq 0$). This means that $z^2 + 1$ has no real roots. But it does have complex roots, namely $\pm i$: substituting $z = i$ yields $i^2+1 = -1+1 = 0$, and similarly for $z = -i$.

In fact, for every complex number c, the quadratic polynomial z^2-c has complex roots; this is the same thing as saying that there are complex numbers w such that $w^2 = c$. That is: every complex number has complex square roots. The next exercise extends this to nth roots and identifies the roots explicitly.

Exercise 1.5. In this exercise, you will show that every nonzero complex number has two distinct square roots, and in fact n distinct nth roots for every $n \geq 1$.

(a) (Warm-up) Find two distinct complex numbers $w_1 \neq w_2$, such that $w_1^2 = w_2^2 = 1 + i$. Hint: use polar coordinates.

(b) Consider the complex number c with polar coordinates (r, θ), where $r > 0$ and $-\pi < \theta \leq \pi$. Define the following two complex numbers, expressed using polar coordinates:

$$\begin{aligned} w_1 &: \left(\sqrt{r}, \frac{\theta}{2}\right) \\ w_2 &: \left(\sqrt{r}, \frac{\theta}{2} + \pi\right). \end{aligned}$$

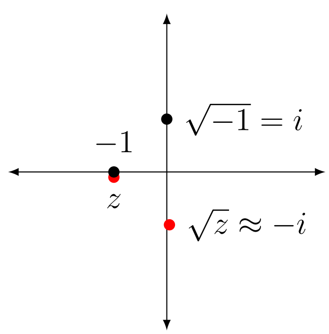
Show that $w_2 = -w_1$ and that $w_1^2 = w_2^2 = c$. Draw a nice picture in the case where $c = i$.

(c) Now fix an integer $n > 2$ and show that every nonzero complex number c has exactly n distinct nth roots. Draw a nice picture of the 6 distinct 6th roots of 1. Hint: use polar coordinates as in part (b).

Remark 1.15. We need to be careful with the square root notation for complex numbers. To see why, recall the conventional notation for square roots of positive real numbers $a > 0$: we denote the positive square root of a by the symbol $\sqrt{a}$, and then the negative square root of a by $-\sqrt{a}$. There is no ambiguity here, because every positive real number has two distinct square roots, one positive and one negative. In the previous exercise, you showed that every nonzero complex number z also has two distinct square roots w_1 and w_2, and that $w_2 = -w_1$. But now there is no consistent way to choose one of these roots as "the positive one" that deserves the label $\sqrt{z}$. This is a bit subtle, but important, so we will elaborate further.

For the purposes of this discussion, let's agree to always use arguments θ from the half-open interval $(-\pi, \pi]$. Furthermore, let's make the *convention* that we will write $\sqrt{z}$ for whichever square root w_1 or w_2 has argument in the half-open interval $(-\pi/2, \pi/2]$. Note that this generalizes the notation for square roots of positive real numbers a, where $\sqrt{a}$ has argument 0 and $-\sqrt{a}$ has argument π. This is a fine and useful convention, but it has some consequences that may surprise you. We list two:

(1) With our convention, we have $\sqrt{-1} = i$ and $\sqrt{-4} = 2i$. Also, we have $\sqrt{4} = 2$. But then even though $(-1)(-4) = 4$, we have $\sqrt{-1}\sqrt{-4} = i \cdot 2i = 2i^2 = -2 = -\sqrt{4}$. This shows that, in general, $\sqrt{zw} \neq \sqrt{z}\sqrt{w}$.



(2) Consider $z = -1 - 0.1i$, which lies close to -1 but just below the real axis (see picture above). Then the argument of z is very close to $-\pi$, so by our convention $\sqrt{z}$ has argument close to $-\pi/2$ and magnitude close to 1. This means that $\sqrt{z}$ is close to the number $-i$. But also by our convention, $\sqrt{-1} = i$. Thus, even though z is very close to -1, the square roots $\sqrt{z}$

and $\sqrt{-1}$ are far apart. We will discuss this odd behavior more fully in Example 1.19.

Here is the takeaway: complex numbers have square roots, but "taking the square root" is not a well-behaved and unambiguous algebraic operation. In any particular case, the use of the familiar notation $\sqrt{z}$ depends on establishing a convention (as above), and must be done carefully.

Now consider a general quadratic polynomial with complex coefficients: $p(z) = az^2 + bz + c$ with $a \neq 0$. The quadratic formula shows that $p(z)$ has roots in the complex numbers, given by the familiar expressions:

$$w_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad w_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

The numbers w_1 and w_2 are distinct roots of $p(z)$, unless the quantity $b^2 - 4ac = 0$, in which case the two roots are equal. So: all complex quadratic polynomials have roots in the complex numbers.

These facts may not be very surprising to you, since you have probably encountered situations in which the quadratic formula confronts you with the need to take the square root of a negative number. What you may not know, however, is that the existence of complex roots does not depend on the polynomial being quadratic.

THEOREM 1.16 (Fundamental Theorem of Algebra). *Consider a polynomial of degree $n \geq 1$ with complex number coefficients:*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \qquad a_n \neq 0.$$

Then there is a complex number c such that $p(c) = 0$.

We will not describe a proof of the Fundamental Theorem, as all proofs (of which there are now many) require non-algebraic ideas coming from subjects such as analysis and topology. In fact, although the statement of this result has been known and used since the 17th century, the first complete proofs were only given in the 19th century.

However, we can use the Fundamental Theorem to derive a corollary that tells us a lot about the structure of polynomials.

COROLLARY 1.17 (Unique Factorization). *Consider a polynomial of degree $n \geq 1$ with complex number coefficients:*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \qquad a_n \neq 0.$$

Then $p(z)$ factors into a product of linear polynomials as follows:

$$p(z) = a_n (z - c_1)(z - c_2) \cdots (z - c_n).$$

The complex roots $c_1, c_2, \ldots, c_n$ are not necessarily distinct, but they are uniquely determined (up to reordering) by the polynomial $p(z)$. In particular, the polynomial $p(z)$ has at most n distinct complex roots.

PROOF. See the optional Section 1.4. □

You should think about this corollary as providing you with a new way of describing polynomials. For instance, suppose you are thinking about a particular cubic polynomial, and you want to tell me which one it is. You could simply say that you are thinking about the polynomial

$$p(z) = z^3 - (11 + 7i)z^2 + (25 + 42i)z - 15 - 35i,$$

listing the coefficients $a_3 = 1$, $a_2 = -11 - 7i$, $a_1 = 25 + 42i$, and $a_0 = -15 - 35i$. On the other hand, you could instead tell me that you are thinking about the polynomial with leading coefficient $a_3 = 1$ and roots $c_1 = 1, c_2 = 5 + 2i, c_3 = 5 + 5i$, in which case I would be thinking about the following expression:

$$(z - 1)(z - 5 - 2i)(z - 5 - 5i).$$

EXERCISE 1.6. Expand this product and verify that that you get the same polynomial $p(z)$.

A few points to mention before we move on: if you give me the second description (leading coefficient a_n and the roots c_i), then I can easily find the first description (all the coefficients a_i) by expanding and collecting like terms—this is what you did in the previous exercise. But the reverse direction isn't so easy: for 2nd degree polynomials the quadratic formula does the trick, and there are similar (although more complicated) formulas for finding the roots of 3rd and 4th degree polynomials. But it is a surprising fact that for 5th and higher-degree polynomials, there are no such elementary formulas (involving only addition, subtraction, multiplication, division, and the extraction of nth roots)! And here we don't mean that no one has yet found them, but rather that it is a proven mathematical truth that no such formulas exist. If you are interested in learning more, keep your eye out for a Galois Theory course later in your mathematical education; this story and its consequences are some of the gems of modern mathematics.

Despite the difficulty of finding the roots of a polynomial, it is useful to know what they are. Think back to the curve-sketching you did in

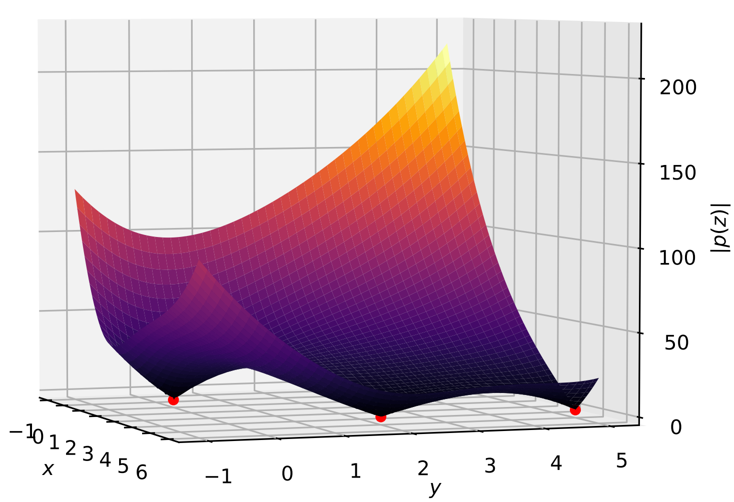


FIGURE 1.3. Graph of the magnitude $|p(z)|$ for the cubic polynomial $p(z) = z^3 - (11+7i)z^2 + (25+42i)z - 15 - 35i$, with roots $c_1 = 1, c_2 = 5 + 2i$, and $c_3 = 5 + 5i$ shown as red dots.

your first calculus course: the real roots of a real polynomial tell you where its graph crosses the x-axis, and thus help you understand its shape. Graphs of complex polynomials live in 4 dimensions (2 for the input variable z and 2 for the output values $p(z)$), so they are difficult to visualize directly. But we can instead think about the graph of the magnitude of the polynomial $|p(z)|$, which is always a real number (in fact nonnegative), and hence only requires 1 dimension. So the graph of $|p(z)|$ lives in 3 dimensions, with the horizontal plane representing the complex input values $z = x + iy$, and the vertical axis the real number outputs $|p(z)|$. The magnitude $|p(c)|$ is zero exactly when the complex number $p(c)$ is itself zero, and so the places where the graph touches the horizontal plane are the roots of the polynomial $p(z)$. Figure 1.3 shows this graph for the cubic polynomial discussed above.

The Fundamental Theorem of Algebra is our first example of the complex numbers being "better" than the real numbers: there are real polynomials without any real roots (such as $z^2 + 1$), but all complex polynomials have complex roots. Moreover, our corollary states that all complex polynomials factor uniquely into a product of linear factors corresponding to the roots, and this factorization gives a good picture of the structure of the polynomial (for a proof, see the optional Section 1.4). In Section 2.1, we provide a second advertisement for the complex numbers involving some beautiful pictures coming from quadratic polynomials, leading ultimately to the topic of infinite sequences, one of the main subjects of this course.

Key points for Section 1.2:

- Existence and description of square roots (Exercise 1.5)
- Statement of the Fundamental Theorem of Algebra (Theorem 1.16)
- Statement of Corollary 1.17 describing the unique factorization of a complex polynomial in terms of its roots

1.3. Complex Functions

Throughout this course, we will think of complex polynomials as the basic type of "nice" function, much as you likely thought about real polynomials as the basic type of "nice" function in your first calculus course. As a high point, in Chapter 4 we will use polynomials to approximate more general complex functions. In this section, we set the stage by talking about the idea of a complex function and investigating some basic examples.

To begin, recall the notion of a real function $f\colon D \to \mathbb{R}$ from one-variable calculus, where the subset $D \subseteq \mathbb{R}$ is the domain of f. At this level of generality, the function f is simply a rule that assigns to every real number x in D another real number $f(x)$. In practice, we often deal with continuous or differentiable functions defined by explicit formulas. A convenient way of displaying a function is to draw its graph, with the domain on the horizontal axis, and the output values on the vertical axis. Figure 1.4 shows the graphs of three familiar functions.

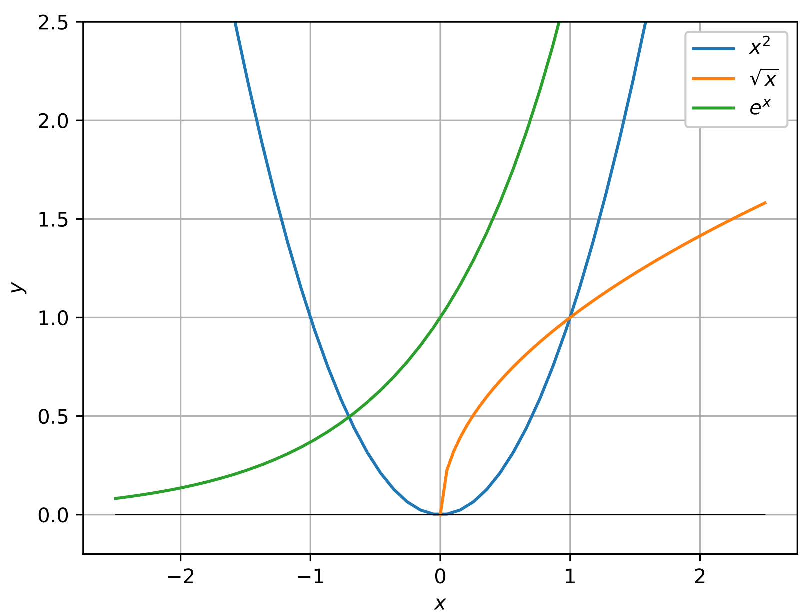
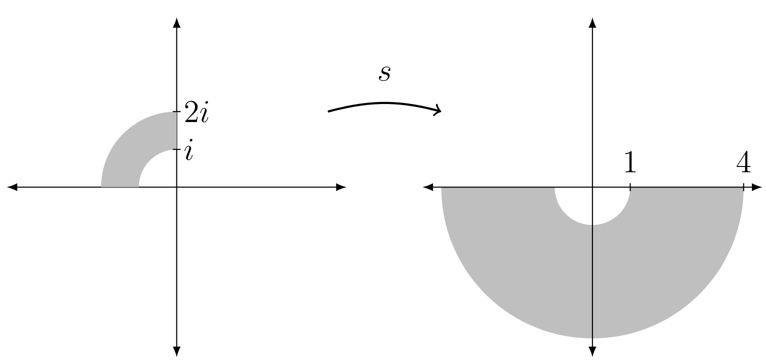


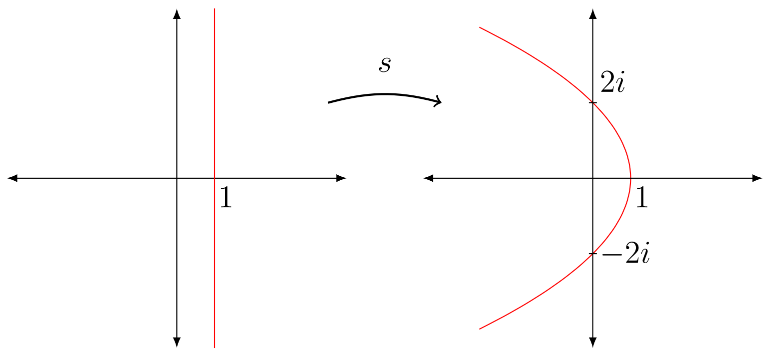
FIGURE 1.4. Graphs of three real functions

Now we turn our attention to complex functions $f\colon D \to \mathbb{C}$, where the domain $D \subseteq \mathbb{C}$ is a subset of the complex plane. Again, such a function f is just a rule that assigns to every complex number z in D another complex number $f(z)$. Just as for real functions, in practice we will deal with nice functions that are defined by explicit formulas—but one of the main goals of this course is to expand our notion of what counts as a "formula." Looking way ahead, in Chapter 4 we will introduce *power series* as formulas that generalize polynomial expressions and are able to describe a large class of important and useful functions.

As mentioned previously during our discussion of complex polynomials, we are not able to directly visualize the graph of a complex function, because it lives in 4-dimensional space: two dimensions are required for the domain, and another two dimensions for the complex output values. So to visualize the behavior of complex functions, we instead draw two copies of the complex plane side-by-side, the first representing the domain of inputs, and the second representing the corresponding outputs.

Example 1.18. Let's see how this works for the squaring function $s(z) = z^2$, with domain $\mathbb{C}$, the entire complex plane. To square a complex number, we double its argument and square its magnitude. The pictures below shows the effect of squaring on (1) an annular region in the 2nd quadrant of the complex plane and (2) the vertical line $x = 1$. Problem 1.13 at the end of the chapter asks you to investigate some other subsets of the complex plane.





EXERCISE 1.7. In the picture above, let $u+iv = s(x+iy)$ denote the squaring function, so that the real variables x, y describe the coordinate axes on the left hand side (inputs), while the real variables u, v describe the coordinate axes on the right hand side (outputs). Verify that the squaring function sends the vertical line $x = 1$ to the sideways parabola defined by $u = 1 - \frac{1}{4}v^2$.

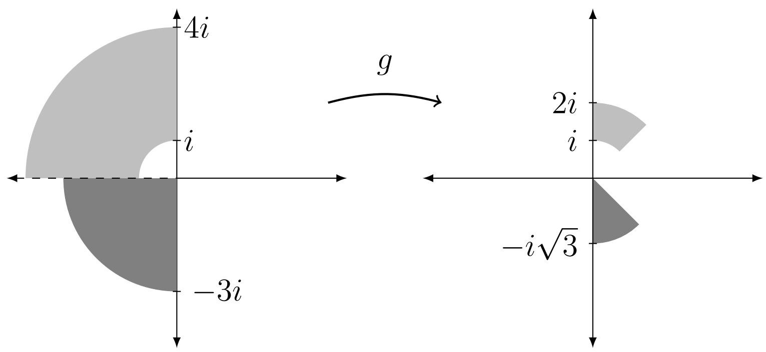
EXAMPLE 1.19. We now wish to define and study a square root function. Before reading further, you should revisit EXERCISE 1.5 and especially Remark 1.15 to appreciate some pitfalls. To deal with these subtleties, we will choose the domain D of our function carefully:

$$D = \mathbb{C} \setminus (-\infty, 0].$$

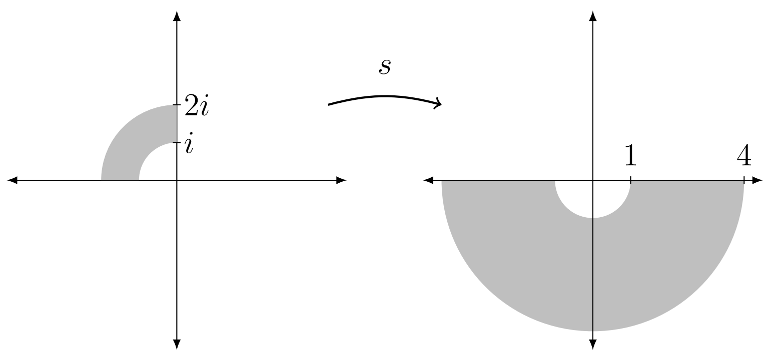
In words, D is obtained from the complex plane $\mathbb{C}$ by removing the nonpositive real axis. Note that every number w in D may be written uniquely in polar form as $w = r(\cos(\theta) + i\sin(\theta))$ where $r > 0$ and $-\pi < \theta < \pi$. Then define

$$g(w) = \sqrt{r}(\cos(\theta/2) + i\sin(\theta/2)).$$

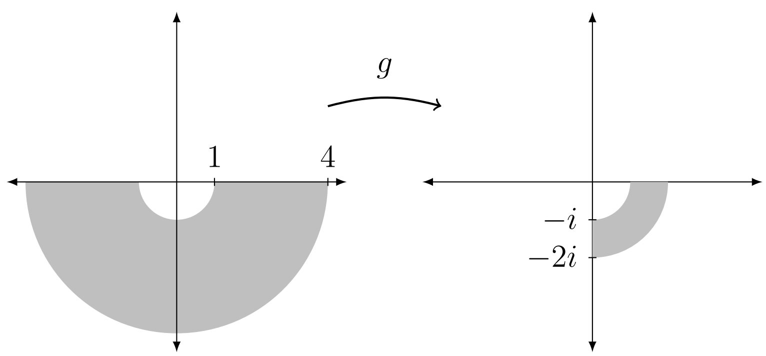
The picture below shows the effect of g on two regions in the 2nd and 3rd quadrants of the complex plane (recall that the negative real axis is not in the domain D of the function):



Now let's revisit the first picture from Example 1.18, the squaring function:



The next picture shows the effect of g on the output region:

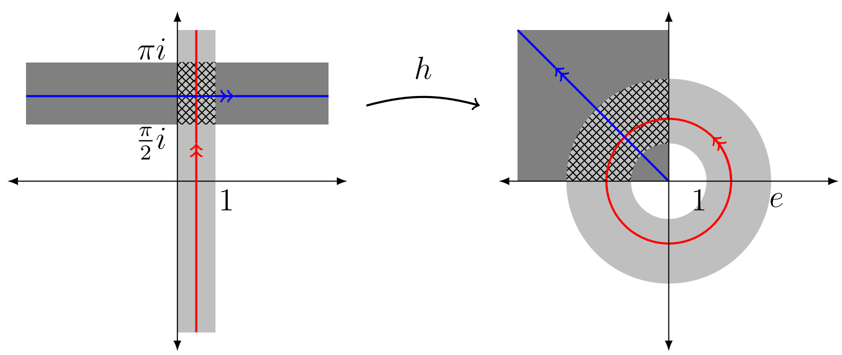


Note the interesting fact that we do not recover the original annular region in the 2nd quadrant, but rather its rotation by π in the 4th quadrant. So, despite the fact that $g(w)$ is a square root of w for every input w, it is not true that g and the squaring function s from Example 1.18 are inverses. But we can fix this if we are more careful about the domain of the squaring function. First note that all of the outputs of g lie in the right half-plane U consisting of complex numbers with positive real part. If we restrict the squaring function to this domain by defining $f\colon U \to \mathbb{C}$ by $f(z) = z^2$, then the functions f and g are inverses of each other: $f(g(w)) = w$ for all w in D and $g(f(z)) = z$ for all z in U. In words: the squaring function f takes the right side of the plane and stretches it open like a fan to cover all of D, the complex plane with the negative real axis removed. The function g undoes this stretching, partially closing the fan to recover the half-plane.

Example 1.20. As a final example of a complex function, consider $h\colon \mathbb{C} \to \mathbb{C}$ defined by the formula

$$h(x + iy) = e^x(\cos(y) + i \sin(y)).$$

Observe the following, and refer to the picture below, which shows the effect of h on both a horizontal and a vertical strip:



- If we set $y = 0$ and consider only real inputs x, then h becomes the ordinary real exponential function e^x. In this way, the complex function h is an *extension* of the real exponential function to the complex plane.
- The function h takes the vertical line $x = a$ to the circle of radius e^a. In fact, as y increases, the function wraps the vertical line counter-clockwise around the circle infinitely many times. For instance, all numbers of the form $a + (2\pi n)i$ are sent by h to the point e^a on the positive real axis.
- The right half plane is sent to the exterior of the unit circle (think of the right half plane as a collection of vertical lines,

and use the previous observation). Likewise, the left half plane is sent to the interior of the unit circle.

- The horizontal line $y = b$ is sent to the ray from the origin at an angle of b radians from the positive real axis. The ray is traced out via an exponential parametrization, approaching the origin as $x \to -\infty$ and traveling away from the origin as $x \to +\infty$.
- The horizontal strip consisting of all complex numbers $x + iy$ with $0 \leq y < 2\pi$ is spread out like a fan to cover the entire complex plane with the exception of the origin; the left half of the strip fills out the interior of the unit circle, while the right half of the strip fills out the exterior. (Think of the horizontal strip as a collection of horizontal lines, and use the previous observation).

REMARK 1.21. The previous three examples provide complex versions of the real functions displayed in Figure 1.4. In general, this will be a major question for us: starting with a familiar real function $f(x)$, is it possible to find a nice complex function $F(z)$ with the property that F agrees with f for real inputs x? This was easy to accomplish for the squaring function, and the complex square root function only required some care to avoid the pitfalls discussed in Remark 1.15. But the final example is surprising: why should we extend the familiar real exponential function e^x by using the cosine and sine functions in just this way? In Chapter 4 we will discover a surprising answer to this question, namely that this is the *only* reasonable way to extend the real exponential to a complex function! We will also be able to find complex versions of many other familiar functions, such as $\arctan(x)$, the inverse of the tangent function. Before we can do any of this, however, we need to study infinite sequences and series—these are the topics of the next two chapters.

> Key points for Section 1.3:
>
> - Idea of a complex function $f\colon D \to \mathbb{C}$ (page 22)
> - Geometry of squaring function (Example 1.18)), square root function (Example 1.19), and complex extension of exponential function (Example 1.20))

1.4. Optional: Polynomial Division and Unique Factorization

We first recall the process of *polynomial division with remainder* by means of an example using real numbers.

Example 1.22. Consider the two polynomials

$$\begin{aligned} p(x) &= x^3 + 2x + 1 \\ d(x) &= x + 1. \end{aligned}$$

We wish to divide $p(x)$ by $d(x)$, leaving a remainder. The idea is to multiply $d(x)$ by a monomial in order to match the highest-order term of $p(x)$, then subtract and repeat until we are left with a polynomial of degree less than the degree of $d(x)$:

$$\begin{array}{r|ccccccc}
 & x^2 & - & x & + & 3 & & \\
\hline
x+1\,| & x^3 & + & & & 2x & + & 1 \\
- & (x^3 & + & x^2) & & & & \\
\hline
 & & & -x^2 & + & 2x & + & 1 \\
- & & & (-x^2 & - & x) & & \\
\hline
 & & & & & 3x & + & 1 \\
- & & & & & (3x & + & 3) \\
\hline
 & & & & & & & -2
\end{array}$$

The upshot of this computation is that we have found a *remainder* polynomial $r(x)$ and a *quotient* polynomial $q(x)$ such that

$$p(x) = d(x)q(x) + r(x).$$

In this example, we have

$$\begin{aligned} r(x) &= -2 \\ q(x) &= x^2 - x + 3. \end{aligned}$$

Indeed, we explicitly check that

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

EXAMPLE 1.23. Here is a more substantial example involving complex numbers:

$$\begin{aligned} p(z) &= 2z^5 + iz^4 + z^3 - 3z^2 + (1-i)z + 1 \\ d(z) &= z^3 - (1+i)z^2 + 1. \end{aligned}$$

We proceed in the same way as above, multiplying $d(z)$ by a monomial in order to match the highest-order term of $p(z)$, then subtracting and repeating:

$$\begin{array}{l|llllll}
 & & & 2z^2 & +(2+3i)z & +5i & \\
\hline
z^3-(1+i)z^2+1 & 2z^5 & +iz^4 & +z^3 & -3z^2 & +(1-i)z & +1 \\
 & -(2z^5 & -(2+2i)z^4 & +0 & +2z^2) & & \\
\hline
 & & (2+3i)z^4 & +z^3 & -5z^2 & +(1-i)z & +1 \\
 & - & ((2+3i)z^4 & +(1-5i)z^3 & +0 & +(2+3i)z) & \\
\hline
 & & & 5iz^3 & -5z^2 & -(1+4i)z & +1 \\
 & - & & (5iz^3 & +(5-5i)z^2 & +0 & +5i) \\
\hline
 & & & & (-10+5i)z^2 & -(1+4i)z & +(1-5i)
\end{array}$$

In this example, the remainder and quotient are

$$\begin{aligned} r(z) &= (-10+5i)z^2 - (1+4i)z + (1-5i) \\ q(z) &= 2z^2 + (2+3i)z + 5i. \end{aligned}$$

EXERCISE 1.8. Check explicitly that $p(z) = d(z)q(z) + r(z)$.

The procedure illustrated in these examples will work for any pair of polynomials $p(z)$ and $d(z)$, yielding a quotient polynomial $q(z)$ and a remainder $r(z)$ of degree strictly less than the degree of $d(z)$. In the case where $r(z) = 0$, we have $p(z) = d(z)q(z)$, and we say that $d(z)$ *evenly divides* $p(z)$, and that $d(z)$ is a *factor* of $p(z)$. In the case where $d(z)$ has degree 1, we say that it is a *linear factor* of $p(z)$.

Linear factors of a polynomial are extremely important, because they correspond to the roots. We prove this carefully in the next proposition.

PROPOSITION 1.24. *Suppose that $p(z)$ is a complex polynomial and c a complex number. Then division by the linear polynomial $z-c$ yields the constant remainder $p(c)$:*

$$p(z) = (z-c)q(z) + p(c)$$

for some polynomial $q(z)$. In particular, $p(c) = 0$ (that is, c is a root of p) if and only if $z - c$ is a factor of $p(z)$.

PROOF. Perform division with remainder of $p(z)$ by the linear polynomial $d(z) = z-c$. Since the remainder must have degree strictly less than the degree of the linear polynomial d, we see that the remainder is actually a constant complex number r (a polynomial of degree 0).

$$p(z) = (z-c)q(z) + r.$$

Now substitute $z = c$. The first term on the right-hand-side becomes zero, and we find that

$$p(c) = r.$$

It follows that the remainder $r = p(c) = 0$ if and only if $d(z) = z - c$ is a factor of $p(z)$. □

PROOF OF COROLLARY 1.17. The corollary has two parts: an *existence* part stating that every polynomial factors as a product of linear polynomials, and a *uniqueness* part stating that the list of roots $c_1, c_2, \ldots, c_n$ is unique up to reordering. We will prove the existence part by contradiction, using a technique sometimes known as a "minimal criminal" argument. Indeed, if the existence statement is not true, then there must be some polynomials $p(z)$ that cannot be factored into a product of linear polynomials—these are the criminals. Among all of these criminal polynomials, let's focus attention on one of least degree

$n \geq 2$—this is our minimal criminal, and we call it $p(z)$. (Observe that $n \geq 2$, since if $p(z)$ had degree 1, it would itself be a linear polynomial $az + b$, and hence not a criminal).

Now apply the Fundamental Theorem of Algebra to $p(z)$, which yields a complex root c such that $p(c) = 0$. By the previous proposition, $z - c$ is a linear factor of $p(z)$, and it evenly divides $p(z)$:

$$p(z) = (z - c)q(z).$$

Note that the degree of $q(z)$ is one less than the degree of $p(z)$, and hence $q(z)$ cannot be a criminal, which means that

$$q(z) = a(z - c_1)(z - c_2) \ldots (z - c_{n-1})$$

for some complex numbers a and c_i. But then substituting into our expression for $p(z)$ reveals that $p(z)$ isn't a criminal after all:

$$p(z) = a(z - c_1)(z - c_2) \cdots (z - c_{n-1})(z - c).$$

This contradiction shows that there are, in fact, no criminals, so every polynomial can be factored as a product of linear factors.

For the uniqueness part, suppose that we have a single polynomial $p(z)$ that can be factored in two possibly different ways:

$$\begin{aligned} p(z) &= a(z - c_1)(z - c_2) \cdots (z - c_n) \\ p(z) &= b(z - d_1)(z - d_2) \cdots (z - d_m). \end{aligned}$$

Here, $c_1, c_2, \ldots, c_n$ is a list of complex roots of $p(z)$, possibly with repetitions, and similarly for $d_1, d_2, \ldots, d_m$. First, some easy observations: the total degree of the polynomial $p(z)$ is the number of linear factors on the right hand side, so it follows that $n = m$, and the two lists of roots c_i and d_i must have the same length. Second, if we were to expand the first expression, the number a would emerge as the coefficient of the leading term z^n. For the same reason, b must be the coefficent of the leading term, so $a = b$. Finally, each number in the list c_i must appear at least once in the list d_i and vice-versa. Indeed, the first expression for $p(z)$ makes it clear that the c_i are the *only* roots of $p(z)$, since plugging in any other number for z would yield a product of nonzero complex numbers, which can't be zero. Similarly, the d_i are the only roots of $p(z)$.

All that remains is to show that the two lists of roots c_i and d_i have the same repetitions. Since we don't care about the ordering of these lists, we need to show that if c_1 appears exactly k times in the first list, then it also appears exactly k times in the second list. So suppose that c_1 appears k times in the first list of roots, and $\ell \geq k$ times in the second list. Our two expressions for $p(z)$ become

$$\begin{aligned} p(z) &= a(z-c_1)^k(z-c_{k+1})\cdots(z-c_n) \\ p(z) &= a(z-c_1)^\ell(z-d_{\ell+1})\cdots(z-d_n) \end{aligned}$$

where the roots $c_{k+1}, \ldots, c_n$ and $d_{\ell+1}, \ldots, d_n$ are distinct from c_1. It follows that $a(z-c_1)^k$ is a factor of $p(z)$, and the first expression allows us to write the quotient as $q(z) = (z-c_{k+1})\cdots(z-c_n)$. In particular, $q(c_1) \neq 0$. But the second expression allows us to write the same quotient as

$$q(z) = (z-c_1)^{\ell-k}(z-d_{\ell+1})\cdots(z-d_n).$$

Note that, if $\ell > k$, then $(z-c_1)$ would be a linear factor of $q(z)$, and $q(c_1) = 0$. Hence, we see that $\ell = k$ as required. □

Key points for Section 1.4:

- Polynomial division with remainder (Examples 1.22 and 1.23)
- Relationship between roots and linear factors (Proposition 1.24)

1.5. In-text Exercises

This section collects the in-text exercises that you should have worked on while reading the chapter.

EXERCISE 1.1 When does equality occur in the statement of the triangle inequality?

EXERCISE 1.2 Recall the algebraic formula for the product of two complex numbers $a + bi$ and $c + di$:

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

(a) Express two generic complex numbers z and w in polar form, and then use the formula above to compute the product zw.
(b) Check that the $\arg(zw) = \arg(z) + \arg(w)$. You will need to make use of the trigonometric identities

$$\begin{aligned} \cos(\theta + \phi) &= \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi) \\ \sin(\theta + \phi) &= \cos(\theta)\sin(\phi) + \sin(\theta)\cos(\phi). \end{aligned}$$

(c) Finally, check that $|zw| = |z||w|$.

EXERCISE 1.3 Can you convince yourself that every nonzero complex number z has exactly one inverse, i.e. that inverses are unique?

EXERCISE 1.4 Show that $z\overline{z} = |z|^2$, so that $z^{-1} = \overline{z}/|z|^2$.

EXERCISE 1.5 In this exercise, you will show that every nonzero complex number has two distinct square roots, and in fact n distinct nth roots for every $n \geq 1$.

(a) (Warm-up) Find two distinct complex numbers $w_1 \neq w_2$, such that $w_1^2 = w_2^2 = 1 + i$. Hint: use polar coordinates.
(b) Consider the complex number c with polar coordinates (r, θ), where $r > 0$ and $-\pi < \theta \leq \pi$. Define the following two complex numbers,

expressed using polar coordinates:

$$\begin{aligned} w_1 &: \left(\sqrt{r}, \frac{\theta}{2}\right) \\ w_2 &: \left(\sqrt{r}, \frac{\theta}{2} + \pi\right). \end{aligned}$$

Show that $w_2 = -w_1$ and that $w_1^2 = w_2^2 = c$. Draw a nice picture in the case where $c = i$.

(c) Now fix an integer $n > 2$ and show that every nonzero complex number c has exactly n distinct nth roots. Draw a nice picture of the 6 distinct 6th roots of 1. Hint: use polar coordinates as in part (b).

EXERCISE 1.6 Expand the following product and verify that that you get the polynomial $p(z)$ on page 18.

$$(z - 1)(z - 5 - 2i)(z - 5 - 5i).$$

EXERCISE 1.7 In the picture above, let $u + iv = s(x + iy)$ denote the squaring function, so that the real variables x, y describe the coordinate axes on the left hand side (inputs), while the real variables u, v describe the coordinate axes on the right hand side (outputs). Verify that the squaring function sends the vertical line $x = 1$ to the sideways parabola defined by $u = 1 - \frac{1}{4}v^2$.

EXERCISE 1.8 Check explicitly that $p(z) = d(z)q(z) + r(z)$ for the polynomials in Example 1.23.

1.6. Problems

1.1. For each of the following pairs of complex numbers z and w, compute the numbers $z + w$, $z - w$, zw, w^{-1}, and z/w. Express all of your answers in cartesian form $a + bi$.

(a) $z = 1 + i, \quad w = \sqrt{3} - i$
(b) $z = -2 - 3i, \quad w = \frac{1}{3}i$
(c) $z = 3, \quad w = -1 + 2i$
(d) $z = -4 - 3i, \quad w = -4 + 3i$

1.2. Consider the following pairs of complex numbers:

(a) $z = 1 + i, \quad w = \sqrt{3} - i$
(b) $z = \sqrt{2} + \sqrt{2}i, \quad w = 3i$
(c) $z = -3, \quad w = -2\sqrt{3} + 2i$
(d) $z = -2i, \quad w = \frac{1}{3} + \frac{\sqrt{3}}{3}i$

For each of the pairs z, w of complex numbers listed above, complete the following:

(i) Plot z and w on the complex plane, and then use the parallelogram law to find the location of $z + w$ and $z - w$.
(ii) Find the polar forms of z and w.
(iii) Use geometric reasoning to find the location of w^{-1}.
(iv) Use the rotation-scale interpretation of multiplication to find the location of zw and z/w

1.3. Let z and w be complex numbers. Show that the following properties are true:

(a) $\overline{z + w} = \overline{z} + \overline{w}$
(b) $\overline{zw} = \overline{z} \cdot \overline{w}$
(c) $|z|^2 = z\overline{z}$ (this is Exercise 1.4)
(d) $|\overline{z}| = |z|$

1.4. Let $z = a + bi$ be a complex number. Show that the following properties are true:

(a) $\mathrm{Re}(z) = \dfrac{z + \overline{z}}{2}$

(b) $\operatorname{Im}(z) = \dfrac{z - \overline{z}}{2i}$
(c) $\operatorname{Re}(iz) = -\operatorname{Im}(z)$
(d) $\operatorname{Im}(iz) = \operatorname{Re}(z)$

1.5. Sketch images of the following, shading the described regions.

(a) $|z| < 4$
(b) $|z - i| < 2$
(c) $1 < |z - i + 2| < 2$
(d) $\big|\operatorname{Re}(z)\big| < |z|$
(e) $0 < \operatorname{Im}(z) < \pi$
(f) $-1 < \operatorname{Re}(z) \le 1$
(g) $|z - 1| < |z|$

1.6. Is it true that $\operatorname{Re}(zw) = \operatorname{Re}(z)\operatorname{Re}(w)$ for all complex numbers z and w? If so, show that it is always true. If not, give an example of complex numbers z and w such that $\operatorname{Re}(zw) \neq \operatorname{Re}(z)\operatorname{Re}(w)$.

1.7. This problem provides an algebraic proof of the triangle inequality, Proposition 1.7. Let z and w be complex numbers.

(a) Show that $|z + w|^2 = |z|^2 + z\overline{w} + w\overline{z} + |w|^2$.
(b) Justify the following chain of equalities and inequalities:

$$z\overline{w} + w\overline{z} = z\overline{w} + \overline{(z\overline{w})} = 2\operatorname{Re}(z\overline{w}) \le 2|z||w|\,.$$

(c) Combine parts (a) and (b) to show that $|z + w|^2 \le (|z|+|w|)^2$. Then take the square root of both sides to obtain the triangle inequality.

1.8. Suppose that $z \neq -1$ is a complex number of unit magnitude: $|z| = 1$. Show that $w = \frac{1+z}{|1+z|}$ is a square root of z. (Hint: compute w^2 and use Exercise 1.4).

1.9. For each of the following complex numbers z, find both square roots. Express your answers in cartesian form $a + bi$.

(a) $z = 2i$
(b) $z = -3i$
(c) $z = 1 + i\sqrt{3}$

1.10. Use the quadratic formula to factor the following polynomial into linear factors. Express both roots in cartesian form $a + bi$.

$$p(z) = 2z^2 - (1+i)z - 2i.$$

1.11. Check that $z = 1$ is a root of the following cubic polynomial

$$p(z) = z^3 - iz^2 - z + i.$$

Use polynomial division to factor $p(z)$ as the product of a linear polynomial and a quadratic. Then use the quadratic formula to finish the factorization of $p(z)$ into linear factors.

1.12. Let $s : \mathbb{C} \to \mathbb{C}$ be the squaring function $s(z) = z^2$. Compute the following and write your answers in cartesian form $a + bi$:

(a) $s(3 - i)$
(b) $s(-\sqrt{2} - 2i)$
(c) $s(\frac{1}{2} + \frac{1}{2}i)$
(d) $s(z)$, where z has polar coordinates $(4, \frac{\pi}{4})$
(e) $s(w)$, where w has polar coordinates $(\frac{1}{3}, \frac{5\pi}{3})$

1.13. Draw nice pictures and write a sentence or two explaining the effect of the squaring function $s(z) = z^2$ on the following regions in the complex plane:

(a) the unit square in the 1st quadrant, with corners $0, 1, 1+i$, and i
(b) the unit square in the 3rd quadrant, with corners $0, -1, -1-i$, and $-i$
(c) the left half plane, defined by $\mathrm{Re}(z) < 0$
(d) the exterior of the circle of radius $1/2$
(e) the vertical line $x = a$
(f) the horizontal line $y = b$
(g) the line $x = y$

1.14. Let $t : \mathbb{C} \to \mathbb{C}$ be the cubing function $t(z) = z^3$. Compute the following and write your answers in cartesian form $a + bi$:

(a) $t(3 - i)$
(b) $t(-\sqrt{2} - 2i)$

(c) $t(\frac{1}{2} + \frac{1}{2}i)$
(d) $t(z)$, where z has polar coordinates $(4, \frac{\pi}{4})$
(e) $t(w)$, where w has polar coordinates $(\frac{1}{3}, \frac{5\pi}{3})$

1.15. Draw nice pictures and write a sentence or two explaining the effect of the cubing function $t(z) = z^3$ on the following regions in the complex plane:

(a) the 1st quadrant of the unit disc, defined by $|z| < 1$, $\text{Re}(z) > 0$, and $\text{Im}(z) > 0$
(b) the 2nd quadrant of the complex plane, defined by $\text{Re}(z) < 0$ and $\text{Im}(z) > 0$
(c) the interior of the circle of radius 2

1.16. Let $g : D \to \mathbb{C}$ be the function from Example 1.19. Compute the following and write your answers in polar form $r(\cos(\theta) + i\sin(\theta))$.

(a) $g(-4 + 4i)$
(b) $g(\sqrt{3} - i)$
(c) $g(z)$, where z has polar coordinates $(12, -\frac{\pi}{6})$
(d) $g(w)$, where w has polar coordinates $(\frac{1}{9}, \frac{3\pi}{4})$

1.17. Draw nice pictures and write a sentence or two explaining the effect of the function g from Example 1.19 on the following regions in the complex plane:

(a) the upper half plane, defined by $\text{Im}(z) > 0$
(b) the exterior of the circle of radius 2 in the 4th quadrant, defined by $|z| > 2$, $\text{Re}(z) > 0$, and $\text{Im}(z) < 0$

1.18. Let $h : \mathbb{C} \to \mathbb{C}$ be the function from Example 1.20. Compute the following:

(a) $h(0)$
(b) $h(1 + \pi i)$
(c) $h(-\frac{1}{3} + \frac{1}{3}i)$

1.19. Draw nice pictures and write a sentence or two explaining the effect of the function h from Example 1.20 on the following lines in the complex plane:

(a) the real axis
(b) the imaginary axis
(c) the line $x = y$
(d) the line $x = 2y$
(e) the line $x = -y$

1.20. Using the function $h : \mathbb{C} \to \mathbb{C}$ defined in Example 1.20,

$$h(x + iy) = e^x(\cos(y) + i \sin(y)),$$

show that $\overline{h(x + iy)} = h\left(\overline{x + iy}\right)$.

CHAPTER 2

SEQUENCES

2.1. Exploration: Complex Dynamics

This section is our second advertisement for the complex numbers, and it has a different flavor than our first advertisement (The Fundamental Theorem of Algebra, Theorem 1.16). We begin by describing a simple procedure: select a complex number c, and consider the quadratic polynomial $p(z) = z^2 + c$. Construct a list $(z_0, z_1, z_2, \dots)$ of complex numbers as follows:

$$
\begin{aligned}
z_0 &= 0 \\
z_1 &= p(z_0) = p(0) = c \\
z_2 &= p(z_1) = p(c) = c^2 + c \\
z_3 &= p(z_2) = p(c^2 + c) = (c^2 + c)^2 + c \\
&\vdots \\
z_n &= p(z_{n-1}) = z_{n-1}^2 + c \\
&\vdots
\end{aligned}
$$

Each term in this list is produced by squaring the previous term and then adding c; we refer to this process as *iterating* the polynomial

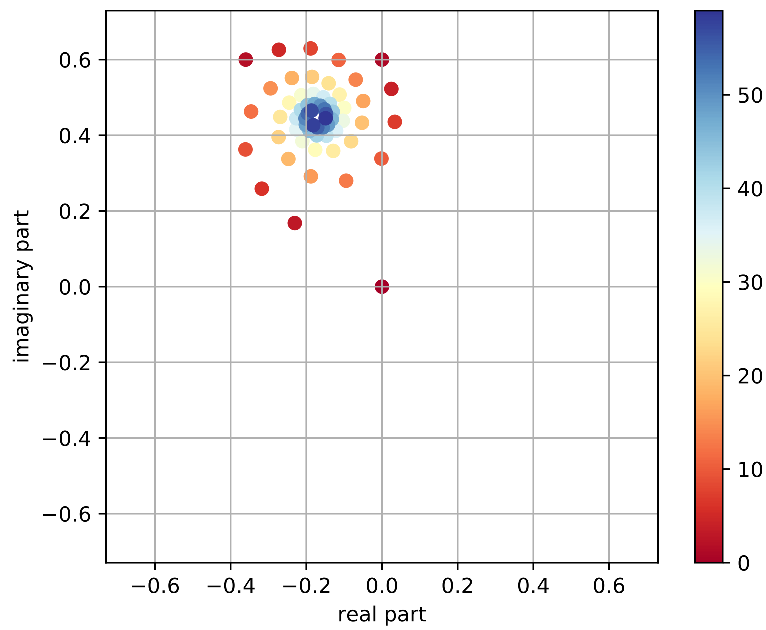


FIGURE 2.1. The first 60 terms of the list obtained by iterating the polynomial $z^2 + 0.6i$ starting with $z_0 = 0$. The colorbar on the right identifies the various terms in the list: the initial terms are red, and the later terms are blue.

$p(z) = z^2 + c$ starting with $z_0 = 0$. Figure 2.1 shows the first 60 terms of the list corresponding to $c = 0.6i$.

EXERCISE 2.1. The term $z_0 = 0$ is represented by the red dot at the origin in Figure 2.1, and the next term $z_1 = c = 0.6i$ is represented by a red dot on the imaginary axis. Thinking purely geometrically (using the rotation-scale interpretation of multiplication and the parallelogram law for addition), find the red dot representing the term $z_2 = c^2 + c$. Can you identify the red dot representing z_3? What about z_4? How far can you go?

A different choice for c would produce a different list of complex numbers, and hence a different picture. In class you will work in groups to investigate these lists as you change c and also as you include more terms. The goal is to identify various behaviors and to discover how those behaviors depend on the choice of c. You will be rewarded with some surprising and beautiful pictures.

2.2. Examples of Sequences

We are now done with the advertisements, and we assume you are enthused about complex numbers. In this section, we begin our study of sequences.

DEFINITION 2.1. A *sequence* is a list of complex numbers in a definite order

$$(z_1, z_2, z_3, \dots),$$

where the dots "…" indicate that the list goes on forever. We also often use the more concise notations $(z_n)_{n\geq 1}$ or simply (z_n) to denote a sequence, where the symbol z_n represents the nth term of the sequence. If all of the numbers in the sequence are actually real numbers, then we say that the sequence is real, and similarly for positive sequences, nonnegative sequences, rational sequences, integer sequences, etc.

REMARK 2.2. Note that in Section 2.1, we started the sequences with the index 0 instead of 1, writing $(z_0, z_1, z_2, \dots)$. Using the concise notation, we would write these sequences as $(z_n)_{n\geq 0}$. The choice of starting point for the index of a sequence is a purely notational matter; it is sometimes more convenient to start at 0 and sometimes more convenient to start at 1 or some other integer. However, once you have chosen notation for a particular sequence, you must stick with it to avoid confusion.

EXAMPLE 2.3. We list several examples of sequences and point out some common misunderstandings:

(a) The counting numbers $(1, 2, 3, \dots)$ form an important sequence of integers. In concise form, we would express this sequence as $(n)_{n\geq 1}$.
(b) The reciprocals of the counting numbers $(1, 1/2, 1/3, \dots)$ forms another important sequence, this time of rational numbers. This sequence $(1/n)$ has a special name: *the harmonic sequence*.

(c) If we alternate the signs in the harmonic sequence, then we obtain the *alternating harmonic sequence*:

$$\left(\frac{(-1)^{n+1}}{n}\right) = \left(1,\ -\frac{1}{2},\ \frac{1}{3},\ -\frac{1}{4},\ \frac{1}{5},\ -\frac{1}{6},\ \dots\right).$$

(d) The list $(i, i, i, \dots) = (i)_{n \geq 1}$, in which every term is equal to the same number i, forms a sequence; such sequences are called *constant*. In particular, sequences may have repetitions.

(e) The list $(1, 2, 3, 1, 2, 3, \dots)$ is a sequence, as is $(3, 2, 1, 3, 2, 1, \dots)$. Despite the fact that each of these sequences contains the same numbers $1, 2, 3$ repeated over and over again, they are not the same sequence: the order of the numbers in a sequence matters.

(f) The prime numbers $(2, 3, 5, 7, 11, 13, 17, 19, 23, \dots)$ form a fascinating sequence of integers. Euclid's *Elements* (c. 300 BCE) contains a proof that there are infinitely many prime numbers, so this truly is an infinite sequence. As of this writing (January, 2019), the largest explicitly known prime number is $2^{82,589,933} - 1$, which has over 24 million base ten digits. This example shows that there are important sequences for which we do not explicitly know all of the terms.

(g) The base ten digits of π provide another interesting example of a sequence for which we do not know all of the terms. This sequence begins as $(3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, \dots)$ and goes on forever. While this sequence is completely fixed and unambiguous, it seems as if it is generated by a random process. Over 22 trillion digits of π have been computed explicitly as of this writing.

(h) In class, you generated the beginnings of many sequences as described in Section 2.1. For each choice of complex number c, you looked at the sequence

$$(0,\ c,\ c^2 + c,\ (c^2 + c)^2 + c,\ \dots).$$

These are examples of *iterative* sequences, as they are produced by iterating a function (in this case the polynomial $z^2 + c$).

EXERCISE 2.2. This exercise asks you to consider some sequences that lie on the unit circle.

(a) Fix a positive integer $m \geq 1$, and consider the complex number $a = \cos(2\pi/m) + i\sin(2\pi/m)$. Describe the sequence consisting of the nonnegative integer powers of a:

$$(a^n) = (1,\, a,\, a^2,\, a^3,\, \dots).$$

Draw a nice picture of this sequence for $m = 6$.

(b) Now fix an irrational real number s in the interval $(0, 1)$, and define the complex number $b = \cos(2\pi s) + i\sin(2\pi s)$. Describe the sequence $(b^n) = (1,\, b,\, b^2,\, b^3,\, \dots)$. What would you say is the key difference between this sequence and the sequence (a^n)?

We now introduce several standard examples of sequences that will appear repeatedly during the course. Our first example generalizes the types of sequences you dealt with in the previous exercise.

EXAMPLE 2.4 (Geometric Sequences). Fix two nonzero complex numbers a and c. The *geometric sequence with initial term a and common ratio c* is

$$(ac^n)_{n\geq 0} = (a, ac, ac^2, ac^3, \dots)$$

Note that we start the indexing with $n = 0$. The nonzero number c is called the *common ratio*, because it is the ratio between each pair of consecutive terms:

$$\frac{ac^{n+1}}{ac^n} = c.$$

For a concrete example, take $a = 2$ and $c = 3i$. Then the corresponding geometric sequence is

$$(2 \cdot (3i)^n) = (2,\, 6i,\, -18,\, -54i,\, 162,\, \dots).$$

REMARK 2.5. In the previous example, we gave an explicit formula for the nth term of a geometric sequence: $z_n = ac^n$. But here is an alternative way to describe the same geometric sequence:

$$z_0 = a \quad \text{and} \quad z_n = c \cdot z_{n-1} \quad \text{for } n \geq 1.$$

This is an example of a *recursive* definition: we explicitly specify one or more initial terms of the sequence (in this case $z_0 = a$), and then we define all later terms z_n via a *recursive formula* involving earlier terms. Most sequences arise recursively, rather than from an explicit formula for the general term. We present several important examples below.

Example 2.6 (Factorial Sequence). Consider the following recursively defined sequence (z_n):

$$z_0 = 1 \quad \text{and} \quad z_n = n \cdot z_{n-1} \quad \text{for } n \geq 1.$$

Here is the beginning of the sequence:

$$(1,\, 1,\, 2,\, 6,\, 24,\, 120,\, \ldots).$$

You may recognize these numbers as the *factorials* of the nonnegative integers:

$$z_n = n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 2 \cdot 1.$$

Note that we are using the convention that $0! = 1$.

Example 2.7 (Inverse Factorial Sequence). The *inverse factorial sequence* is given by the reciprocals of the terms of the factorial sequence:

$$\left(\frac{1}{n!}\right) = \left(1,\, 1,\, \frac{1}{2},\, \frac{1}{6},\, \frac{1}{24},\, \frac{1}{120},\, \ldots\right).$$

Exercise 2.3. Let (w_n) denote the inverse factorial sequence from Example 2.7. Give a recursive definition for the sequence (w_n).

Example 2.8 (Binomial Sequences). This example introduces an important collection of numbers called *binomial coefficients* that arise in a wide variety of mathematical and scientific contexts. We begin by computing powers of the linear polynomial $1 + z$, focusing on the coefficients that appear in the expanded forms:

$$\begin{aligned}(1+z)^0 &= 1\\ (1+z)^1 &= 1+1z\\ (1+z)^2 &= 1+2z+1z^2\\ (1+z)^3 &= 1+3z+3z^2+1z^3\\ &\vdots\end{aligned}$$

In general, when we expand $(1+z)^n$, we will obtain a list of $n+1$ nonzero coefficients that we denote $\binom{n}{k}$ (pronounced "n choose k"):

$$(1+z)^n = \binom{n}{0} + \binom{n}{1}z + \binom{n}{2}z^2 + \cdots + \binom{n}{n-1}z^{n-1} + \binom{n}{n}z^n.$$

To be clear: the symbol $\binom{n}{k}$ denotes the coefficient of z^k in the expansion of $(1+z)^n$. Note that $\binom{n}{k} = 0$ for $k > n$, because z^n is the highest degree term that appears in the expansion of $(1+z)^n$. Thus, for each integer $n \geq 0$, we have a sequence consisting of $n+1$ nonzero terms followed by infinitely many zeros. For instance, here is the sequence of binomial coefficients $\binom{4}{k}$, obtained from the coefficients of $(1+z)^4$:

$$(1, 4, 6, 4, 1, 0, 0, 0, \dots).$$

Later in the course (Example 4.60) we will generalize these binomial sequences from the polynomials $(1+z)^n$ to the more general functions $(1+z)^p$ where p is any real number.

Remark 2.9. We have chosen to introduce the binomial coefficients via the polynomials $(1+z)^n$, because this course is focused on functions. In a discrete mathematics or combinatorics course, these numbers would instead be introduced as the answer to a certain counting problem which explains the phrase "n choose k." The connection with polynomials described above would then become a result called *the binomial theorem*. Problem 2.18 leads you through this combinatorial interpretation of the binomial coefficients.

You may recognize the binomial coefficients $\binom{n}{k}$ from Example 2.8 as the entries of Pascal's triangle (see Figure 2.2). You might even

$n = 0$							1						
$n = 1$						1		1					
$n = 2$					1		2		1				
$n = 3$				1		3		3		1			
$n = 4$			1		4		6		4		1		
$n = 5$		1		5		10		10		5		1	
$n = 6$	1		6		15		20		15		6		1

FIGURE 2.2. The first 7 rows of Pascal's triangle. The kth entry in the nth row is the binomial coefficient $\binom{n}{k}$, obtained by adding the $(k-1)$st and kth entries of row $n-1$.

recall how to generate this triangle: each row begins and ends with a 1, and the interior terms in a row are obtained by adding the two terms just above in the previous row. For instance, in Figure 2.2, the red 15 in the row $n = 6$ is the sum of the red 5 and 10 just above in the row $n = 5$.

This rule for generating Pascal's triangle amounts to saying that the binomial coefficients $\binom{n}{k}$ have a recursive definition: terms in the nth sequence are determined by terms in the $(n-1)$st sequence. For instance, the relationship $15 = 5 + 10$ indicated by the red numbers in Figure 2.2 corresponds to the following equality of binomial coefficients:

$$\binom{6}{2} = \binom{5}{1} + \binom{5}{2}.$$

We make all this precise in the next proposition.

PROPOSITION 2.10. *Let $n \geq 0$ be a positive integer. The binomial coefficients $\binom{n}{k}$ may be described as follows:*

(a) if $k = 0$ or $k = n$, then $\binom{n}{k} = 1$;
(b) if $k > n$, then $\binom{n}{k} = 0$;
(c) if $0 < k < n$, then

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

PROOF. Part (a) follows directly from the expansion of the polynomial $(1+z)^n$: the constant term $\binom{n}{0}$ and the coefficient $\binom{n}{n}$ of z^n

are both 1. Similarly, part (b) follows from the fact that $(1+z)^n$ is a polynomial of degree n, so all higher degree coefficients $\binom{n}{k}$ for $k > n$ are zero.

For part (c), we need to relate the coefficients of $(1+z)^n$ to the coefficients of $(1+z)^{n-1}$. We do this by first factoring $(1+z)^n$ as the product $(1+z)^{n-1}(1+z)$, then expanding $(1+z)^{n-1}$ using binomial coefficients, and finally multiplying by $(1+z)$.

$$\begin{aligned}
(1+z)^n &= (1+z)^{n-1}(1+z) \\
&= \left(\binom{n-1}{0} + \binom{n-1}{1}z + \cdots + \binom{n-1}{n-1}z^{n-1}\right)(1+z) \\
&= \binom{n-1}{0} + \binom{n-1}{1}z + \binom{n-1}{2}z^2 + \cdots + \binom{n-1}{n-1}z^{n-1} \\
&\qquad + \binom{n-1}{0}z + \binom{n-1}{1}z^2 + \cdots + \binom{n-1}{n-1}z^n.
\end{aligned}$$

Now we carefully combine the coefficients of like terms, and we notice a pattern:

$$\begin{aligned}
(1+z)^n &= \binom{n-1}{0} + \left(\binom{n-1}{0} + \binom{n-1}{1}\right)z \\
&\qquad + \left(\binom{n-1}{1} + \binom{n-1}{2}\right)z^2 + \cdots \\
&\qquad + \left(\binom{n-1}{k-1} + \binom{n-1}{k}\right)z^k + \cdots \\
&\qquad + \left(\binom{n-1}{n-2} + \binom{n-1}{n-1}\right)z^{n-1} + \binom{n-1}{n-1}z^n.
\end{aligned}$$

For $0 < k < n$, the coefficient of z^k in this expression is the sum $\binom{n-1}{k-1} + \binom{n-1}{k}$. But this is the expansion of $(1+z)^n$, so the coefficient of z^k must also be $\binom{n}{k}$, which proves the claim:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

□

Remark 2.11. Later in this chapter (page 84 of Section 2.6), we will talk about *the principle of mathematical induction*. Using induction and the recursion from the previous proposition, we will be able to establish an explicit formula for the binomial coefficients in terms of factorials:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

See Proposition 2.45 for the details of the proof.

Key points for Section 2.2:

- Sequences (Definition 2.1)
- Geometric sequences (Example 2.4)
- Recursive sequences (Remark 2.5) and factorials (Example 2.6)
- Binomial sequences (Example 2.8) and Pascal's triangle recursion (Proposition 2.10).

2.3. Boundedness

During the in-class groupwork associated with Section 2.1, you noticed that sequences can either be *bounded* or *unbounded*, and this distinction led to a strange and beautiful picture. We now define these notions carefully.

Definition 2.12. A sequence $(z_1, z_2, z_3, \dots)$ is *bounded* if there is a fixed real number $B > 0$ such that all terms in the sequence have magnitude less than or equal to B: for all indices n, we have $|z_n| \leq B$. The number B is called a *bound* for the sequence. Visually, B is a bound for the sequence if all of the numbers z_n are contained in the disc of radius B centered at the origin of the complex plane (see Figure 2.3). A sequence is *unbounded* if it is not bounded.

This formal definition captures the intuitive notion that for some sequences we can draw a large circle that contains the entire sequence

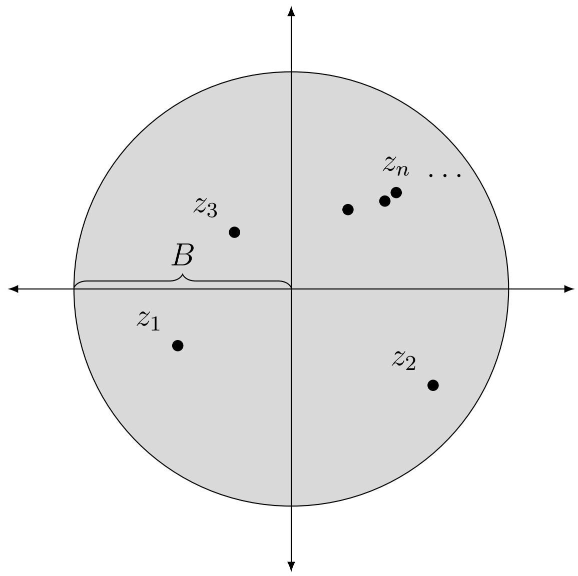


FIGURE 2.3. A bounded sequence is entirely contained within a disc of some finite radius B.

(these are the bounded sequences), whereas other sequences cannot be "fenced in" in this way.

EXERCISE 2.4. Which of the sequences in Example 2.3 are bounded, and which are unbounded?

(a) the counting numbers (n)
(b) the harmonic sequence $(1/n)$
(c) the alternating harmonic sequence $((-1)^{n+1}/n)$
(d) the constant sequence $(i)_{n\geq 1}$
(e) the sequence $(1, 2, 3, 1, 2, 3, \dots)$
(f) the sequence of prime numbers $(2, 3, 5, 7, 11, \dots)$
(g) The digits of π in base ten $(3, 1, 4, 1, 5, 9, \dots)$
(h) The iterative sequences $(0,\ c,\ c^2 + c,\ (c^2 + c)^2 + c,\ \dots)$

In the previous exercise, you probably answered that the sequence of counting numbers $(1, 2, 3, \dots)$ is unbounded. This is correct, and is a fact of such fundamental importance that we record it as a named property:

Archimedean Property of $\mathbb{R}$: The counting numbers $(1, 2, 3, \dots)$ form an unbounded sequence of real numbers. Explicitly, no choice of real number $B > 0$ provides a bound. That is, for every choice of real number $B > 0$, there exists a positive integer $n > 0$ such that $n > B$.

To give a concrete instance of the final sentence above: suppose that I were to propose $B = 1000\pi$ as a bound for the counting numbers. You could immediately show that I am mistaken by exhibiting the integer $n = 4000$, which is bigger than my proposed bound of 1000π. Visually (see Figure 2.3), no finite disc contains all of the counting numbers.

REMARK 2.13. While the Archimedean Property may seem completely obvious, there are in fact important number systems that do not have this property. If you are interested in learning more, perform a web search for *p-adic numbers*.

EXERCISE 2.5. Consider the recursively defined sequence (h_n):

$$h_1 = 1 \quad \text{and} \quad h_n = h_{n-1} + \frac{1}{n} \quad \text{for } n \geq 2.$$

(a) Compute the first 5 terms of the sequence $(h_1, h_2, h_3, h_4, h_5, \dots)$ by hand.

(b) Use a web browser to navigate to SageMathCell, located at

https://sagecell.sagemath.org

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise.

```
N = 100
tail_size = 10
h = 1.0
for n in range(2, N + 1 - tail_size):
    h = h + 1/n
for n in range(N + 1 - tail_size, N + 1):
    h = h + 1/n
    print("h_{:d} = ".format(n) + str(h))
```

Now click Evaluate. This code computes the first $N = 100$ terms of the sequence (h_n) and prints out the last tail_size $= 10$ computed terms to the screen. By changing the values of N and tail_size, you can investigate the behavior of the sequence. Based on your investigations, do you think the sequence (h_n) is bounded?

To conclude this section, we return to the context of Section 2.1, where we formed sequences by iterating the values of a complex polynomial $z^2 + c$, starting with the value $z_0 = 0$.

$\mathbf{c = 1}$: This choice for c yields the sequence $(0, 1, 2, 5, 26, 677, \dots)$. Each term in this sequence is one more than the square of the previous term, so the terms eventually get larger than any fixed real number B. Hence this sequence is unbounded.

$\mathbf{c = 0}$: Now we get the constant sequence $(0, 0, 0, 0, \dots)$. In particular, this sequence is bounded. In fact, any real number $B > 0$ serves as a bound.

$\mathbf{c = -1}$: This yields the sequence $(0, -1, 0, -1, 0, \dots)$ which is bounded, with any number $B \geq 1$ serving as a bound.

$\mathbf{c = 0.5i}$: This choice for c produces the sequence displayed in Figure 2.4. The picture certainly indicates that the sequence is bounded. But something even more interesting is going on. Here are the numerical values of the beginning of the sequence, where we have kept two decimal places of precision:

$$
\begin{aligned}
z_0 &= 0 \\
z_1 &= 0.50i \\
z_2 &= -0.25 + 0.50i \\
z_3 &\approx -0.19 + 0.25i \\
&\vdots \\
z_{14} &\approx -0.12 + 0.40i
\end{aligned}
\qquad
\begin{aligned}
z_{15} &\approx -0.14 + 0.41i \\
&\vdots \\
z_{27} &\approx -0.13 + 0.39i \\
z_{28} &\approx -0.14 + 0.39i \\
z_{29} &\approx -0.14 + 0.39i \\
z_{30} &\approx -0.14 + 0.39i
\end{aligned}
$$

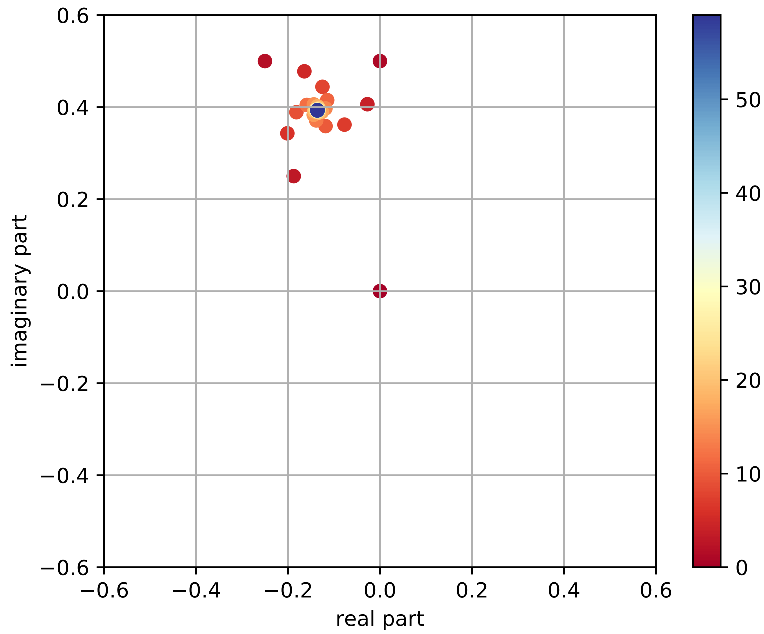


FIGURE 2.4. The first 60 terms of the sequence obtained by iterating the polynomial $z^2+0.5i$ starting with $z_0 = 0$. The colorbar on the right identifies the various terms in the sequence: the initial terms are red, and the later terms are blue.

The sequence appears to settle down to the value $-0.14 + 0.39i$. But if we keep track of more decimal places, we see that this isn't quite correct. Here are terms 30–49 with 4 decimal places of precision:

$$
\begin{aligned}
z_{30} &\approx -0.1353 + 0.3924i & z_{40} &\approx -0.1359 + 0.3930i \\
z_{31} &\approx -0.1357 + 0.3938i & z_{41} &\approx -0.1360 + 0.3932i \\
z_{32} &\approx -0.1367 + 0.3931i & z_{42} &\approx -0.1361 + 0.3931i \\
&\vdots & &\vdots \\
z_{37} &\approx -0.1357 + 0.3931i & z_{47} &\approx -0.1360 + 0.3931i \\
z_{38} &\approx -0.1361 + 0.3933i & z_{48} &\approx -0.1360 + 0.3931i \\
z_{39} &\approx -0.1361 + 0.3930i & z_{49} &\approx -0.1360 + 0.3931i
\end{aligned}
$$

Now it seems that the sequence has settled on $-0.1360 + 0.3931i$. But again, if we look more precisely, we see that the sequence is actually still changing:

$$
\begin{aligned}
z_{44} &\approx -0.135959 + 0.393127i & z_{52} &\approx -0.136026 + 0.393071i \\
z_{45} &\approx -0.136064 + 0.393102i & z_{53} &\approx -0.136002 + 0.393065i \\
z_{46} &\approx -0.136016 + 0.393026i & z_{54} &\approx -0.136003 + 0.393085i \\
z_{47} &\approx -0.135969 + 0.393085i & z_{55} &\approx -0.136019 + 0.393078i \\
z_{48} &\approx -0.136028 + 0.393105i & z_{56} &\approx -0.136009 + 0.393068i \\
z_{49} &\approx -0.136028 + 0.393053i & z_{57} &\approx -0.136004 + 0.393078i \\
z_{50} &\approx -0.135987 + 0.393067i & z_{58} &\approx -0.136014 + 0.393080i \\
z_{51} &\approx -0.136009 + 0.393096i & z_{59} &\approx -0.136012 + 0.393072i
\end{aligned}
$$

While the terms of this sequence are continuing to change, the changes are getting smaller and smaller, and the terms seem to be approaching some particular value that they never quite reach. This is the phenomenon of *convergence*, which is the topic of the next section.

> Key points for Section 2.3:
>
> - Bounded sequences (Definition 2.12)
> - Archimedean property of $\mathbb{R}$ (page 50)

2.4. Convergence

Consider the harmonic sequence $(1/n) = (1, 1/2, 1/3, \dots)$. Intuitively, these numbers are getting closer and closer to zero, but they never quite arrive: each term is smaller than the previous one, but they are all nonzero. On the other hand, it seems that they get arbitrarily close to zero. We want to express this behavior by saying that the sequence *converges* to 0. In order to fully develop this concept, we need to move beyond verbal description and make a formal definition.

As a first step toward a definition, let's focus on the idea of "getting arbitrarily close to zero." The first column below displays positive real numbers $d > 0$, each of which we interpret as a "desired closeness to zero." The second column contains corresponding indices N beyond which all terms of the sequence achieve the desired closeness d: for $n \geq N$, we have $0 < 1/n < d$.

The terms of $(1/n)$ *get arbitrarily close to zero*

$d =$ desired closeness	$N =$ index to achieve d
1	2
0.5	3
0.01	101
0.0034	295
$\vdots$	$\vdots$

Let's check the entries of the table:

- The first line says that if we consider indices $n \geq N = 2$, then we should obtain terms smaller than $d = 1$. And indeed this is true: if $n \geq 2$, then $1/n \leq 1/2 < 1$.
- The second line says that if we consider indices $n \geq 3$, then we should obtain terms smaller than 0.5. And this is also true: if $n \geq 3$, then $1/n \leq 1/3 < 0.5$.
- If we consider indices $n \geq 101$, then we obtain terms smaller than 0.01: if $n \geq 101$, then $1/n \leq 1/101 < 0.01$.
- If $n \geq 295$, then $1/n \leq 1/295 \approx 0.0033898 < 0.0034$, so the fourth line holds as well.

Here is the takeaway from this discussion: no matter what positive number $d > 0$ shows up in the first column, it will be possible to find a valid index N to put in the second column. Indeed, we simply choose any integer $N > 1/d$. Any such N is a valid choice, because if we consider indices $n \geq N$, then $1/n \leq 1/N < d$ as required. (Note that the Archimedean property from page 50 guarantees that there *is* an integer greater than $1/d$.)

We are starting to nail down what we mean by "getting arbitrarily close to zero": no matter what positive number $d > 0$ shows up in the first column above, we can always find a corresponding index N for the second column. To emphasize the fact that the positive number $d > 0$ in the first column can be arbitrary, we imagine playing a game. In this game, there are two players (you and me), and we have different roles: my role is to convince you that the terms of the sequence $(1/n)$ are getting arbitrarily close to zero; your role is to be skeptical.

The Convergence Game (for the harmonic sequence):

(1) You go first, and you challenge me with a small distance $d > 0$; you can choose the positive real number d to be as small as you like, but you have to make a decision and tell me what number you are thinking of.

(2) Now it is my turn. Knowing your choice of distance d, I investigate the terms of the harmonic sequence $(1/n)$ and try to find an index N such that *all* the terms with index N or greater are closer to zero than your distance d. If there is such an index N, then I announce it; if no such N exists, then I lose the game.

(3) Now it is your turn again, and you verify my choice by trying to demonstrate that $1/n < d$ whenever $n \geq N$. If you can find a counterexample to this assertion, then you have shown that my N is not valid and I lose; if your demonstration succeeds, then my N is valid, and I win this round of the game.

(4) Now we return to step (1) and play another round.

To say that the harmonic sequence $(1/n)$ converges to 0 is to say that I will win every round of this game: no matter how small of a distance d you choose, I can specify an index value $N > 1/d$ such that beyond the Nth term of the sequence, all the terms are smaller than d. This is just a rephrasing of the earlier statement that no matter what d shows up in the first column of the table, I can find a valid index N for the second column.

With the convergence game in mind, we now present the formal definition of convergence:

DEFINITION 2.14. A sequence (z_n) *converges* to a complex number w if for all real numbers $d > 0$, there exists an index N such that $|z_n - w| < d$ for all indices $n \geq N$. In this case, we say that w is the *limit* of the sequence and write $\lim_{n\to\infty} z_n = w$. If no such w exists for the sequence (z_n), then we say that the sequence *does not converge* or *diverges.*

This definition has several parts, so we will spend some time unpacking it. Throughout the discussion, keep in mind that it precisely captures the behavior of recent examples, where the terms of a sequence $(z_1, z_2, z_3, \dots)$ are getting arbitrarily close to some number w, but may never actually reach that number. The definition itself is a generalization and distillation of the convergence game described earlier for the harmonic sequence.

- The complex number w is playing the role of zero in the harmonic example: instead of trying to show that the reciprocals $1/n$ are getting arbitrarily close to zero, we are now trying to show that the terms of the given sequence (z_n) are getting arbitrarily close to w. Note that the distance from z_n to w is given by the magnitude $|z_n - w|$.
- The positive real number d is playing the role of the distance you choose in the game; the phrase "for all real numbers $d > 0$" captures the notion that we continue to play rounds of the game indefinitely: in order to assert convergence, I must be able to win every round of the game, no matter what choice you make for the distance d.
- The phrase "there exists an index N such that $|z_n - w| < d$ for all indices $n \geq N$" represents steps (2) and (3) of the game (my turn and your verification), where I respond with a particular index value N, based on your choice of distance d.

Now, instead of verifying that the reciprocals $1/n$ for $n \geq N$ are smaller than d, you must verify that all terms z_n for $n \geq N$ are closer to w than the distance d.

We now state the convergence game for a general sequence (z_n). Definition 2.14 says that if I can win every round of this game, then the sequence (z_n) converges to w:

The Convergence Game (for $\lim_{n\to\infty} z_n = w$):

(1) You go first, and you challenge me with a small distance $d > 0$.
(2) Knowing your choice of distance d, I investigate the terms of the sequence (z_n) and try to find an index N such that *all* the terms with index N or greater are closer to w than your distance d. If there is such an index N, then I announce it; if no such N exists, then I lose the game.
(3) You now verify my choice by trying to demonstrate that the magnitude $|z_n - w| < d$ whenever $n \geq N$. If you can find a counterexample to this assertion, then you have shown that my N is not valid and I lose; if your demonstration succeeds, then my N is valid, and I win this round of the game.
(4) Now we return to step (1) and play another round.

Example 2.15. As an example of playing the convergence game, let's show that the sequence $(z_n) = \left(\frac{n-1}{n+1}\right)$ converges to $w = 1$. You begin by challenging me with a positive real number $d > 0$. I must now find an index N such that $n \geq N$ implies that $|\frac{n-1}{n+1} - 1| < d$. Let's do some preliminary scratch work (this is often the best way to get started). Everything between the two horizontal lines below is scratch work, helping me to discover my choice of N.

Scratch work: Quantity I want to be smaller than d:

$$\left|\frac{n-1}{n+1} - 1\right| = \left|\frac{n-1-(n+1)}{n+1}\right| = \left|\frac{-2}{n+1}\right| = \frac{2}{n+1}.$$

But $2/(n+1) < d$ exactly when $n+1 > 2/d$. I now have an idea for how to choose N, and I can exit the scratchwork.

I am ready to take my turn now, and I choose an integer $N > 2/d$. Now you verify that my choice of N is valid: suppose that $n \geq N$. Then $n + 1 \geq N + 1 > N > 2/d$, and so $2/(n + 1) < d$. But then the distance from the term $z_n = (n - 1)/(n + 1)$ to the proposed limit $w = 1$ is

$$\left|\frac{n-1}{n+1} - 1\right| = \left|\frac{n-1-(n+1)}{n+1}\right| = \left|\frac{-2}{n+1}\right| = \frac{2}{n+1} < d.$$

This argument shows that I can win every round of the convergence game, so $\lim_{n\to\infty} \frac{n-1}{n+1} = 1$.

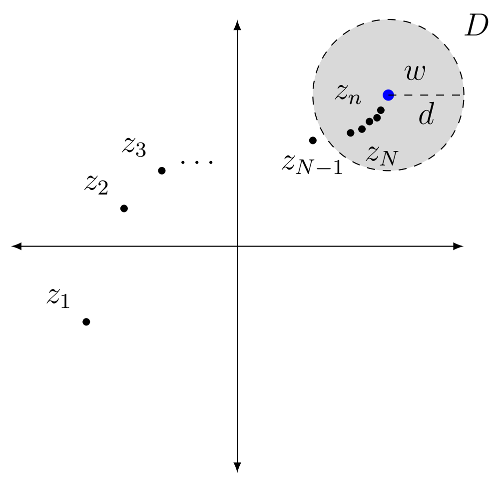


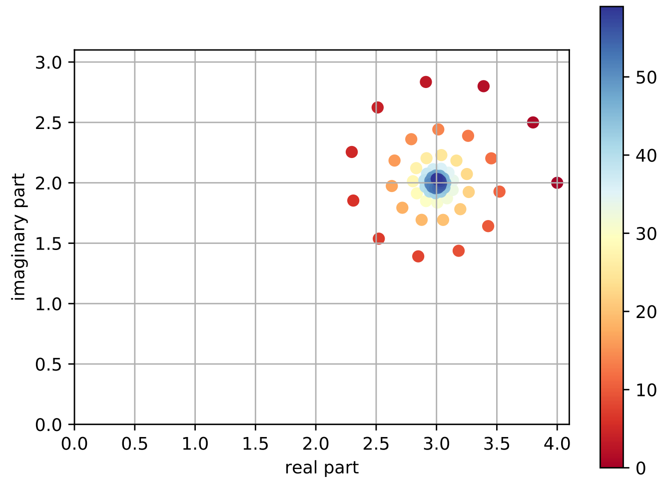
FIGURE 2.5. Convergence

As illustrated in Figure 2.5, it is helpful to think about the definition of convergence geometrically: the distance d determines a disc of radius d centered at the point w in the complex plane. The condition $|z_n - w| < d$ just says that the term z_n is contained in that disc. So: convergence to w can be expressed by saying that for every disc D centered at w, the sequence is *eventually* contained in D: there is an index N such that for all $n \geq N$, the term z_n is in D.

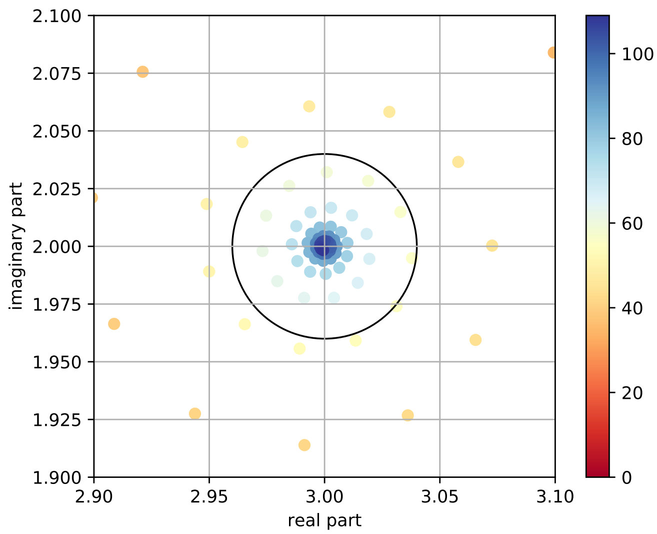
EXAMPLE 2.16. Consider the sequence

$$(z_n) = \left(3 + 2i + (0.8 + 0.5i)^n\right).$$

A picture of this sequence is shown below, suggesting that the sequence converges to $w = 3 + 2i$.



Suppose that you choose the distance $d = 0.04$ for a round of the convergence game. To win this round, I must find an index N such that for all $n \geq N$, the term z_n is contained in the disc of radius $d = 0.04$ centered at $w = 3 + 2i$. The following picture suggests that the index $N = 80$ works:



To actually prove the convergence to w, we play the convergence game. First note that $(z_n) = (w + c^n)$ where $w = 3 + 2i$ is our proposed limit, and $c = 0.8 + 0.5i$. As usual, you go first, choosing a positive distance $d > 0$. I must find a corresponding valid index N. We do some scratchwork:

Scratch work: Quantity I want to be smaller than d:

$$|z_n - w| = |w + c^n - w| = |c^n| = |c|^n .$$

We have

$$|c| = \sqrt{(0.8)^2 + (0.5)^2} \approx 0.94 < 1.$$

But $d > |c|^n$ is equivalent to the condition that

$$\ln(d) > \ln(|c|^n) = n \ln(|c|).$$

Since $0 < |c| < 1$, the logarithm $\ln(|c|) < 0$ is negative, and so the inequality reverses when we divide: $\ln(d)/\ln(|c|) < n$. I now have an idea for how to choose N, and I can exit the scratchwork.

For my turn, I choose an index $N > \ln(d)/\ln(|c|)$. To verify my choice, suppose that $n \geq N$. Since $\ln(|c|) < 0$, we have

$$\ln(|c|^n) = n \ln(|c|) \leq N \ln(|c|) < \ln(d).$$

Applying the exponential function e^x to both sides of this inequality then yields $|c|^n < d$. Finally, we find that

$$|z_n - w| = |w + c^n - w| = |c^n| = |c|^n < d.$$

This shows that I can win every round of the convergence game, and so $\lim_{n\to\infty} z_n = w$.

Example 2.17 (Babylonian sequence for $\sqrt{2}$). Consider the following recursively defined sequence of real numbers (x_n):

$$x_0 = 1 \qquad \text{and} \qquad x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) \quad \text{for } n \geq 1.$$

Here are the first 8 terms of the sequence, to 14 decimal places of precision:

$$
\begin{aligned}
x_0 &= 1 & x_4 &\approx 1.41421356237469 \\
x_1 &= 1.5 & x_5 &\approx 1.41421356237309 \\
x_2 &\approx 1.41666666666667 & x_6 &\approx 1.41421356237309 \\
x_3 &\approx 1.41421568627451 & x_7 &\approx 1.41421356237309
\end{aligned}
$$

It certainly seems that this sequence is converging—and quickly!—to a value close to 1.41421356237309. You may recognize this number as the beginning of the decimal expansion of $\sqrt{2}$. And indeed, here are the squares of the first 8 terms listed above, again to 14 decimal places of precision:

$$
\begin{aligned}
x_0^2 &= 1 & x_4^2 &\approx 2.00000000000451 \\
x_1^2 &= 2.25 & x_5^2 &\approx 2.00000000000000 \\
x_2^2 &\approx 2.00694444444444 & x_6^2 &\approx 2.00000000000000 \\
x_3^2 &\approx 2.00000600730488 & x_7^2 &\approx 2.00000000000000
\end{aligned}
$$

Be careful: despite appearances, none of the squares x_n^2 are exactly equal to 2, but they are getting arbitrarily close as n increases. That is, the sequence of squares (x_n^2) is converging to 2, and the sequence (x_n) itself is converging to $\sqrt{2}$. We will prove this in Example 2.48 of Section 2.6. The next exercise asks you to investigate some similar sequences.

EXERCISE 2.6. Fix a positive real number $r > 0$ and an initial guess $x_{\text{guess}} > 0$ for $\sqrt{r}$. Consider the recursively defined sequence (x_n):

$$x_0 = x_{\text{guess}} \quad \text{and} \quad x_n = \frac{1}{2}\left(x_{n-1} + \frac{r}{x_{n-1}}\right) \quad \text{for } n \geq 1.$$

Use a web browser to navigate to SageMathCell, located at https://sagecell.sagemath.org

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise in the process. Now click Evaluate. This code prints the first $N = 10$ terms of the sequence (x_n) for $r = 17$ and $x_{\text{guess}} = 4$. It then prints out the values of r and x_{N-1}^2. By changing the values of N, r, and x_{guess}, you can investigate the behavior of these sequences. Based on your investigations, do you think that (x_n) always converges to $\sqrt{r}$?

```
r = 17
x_guess = 4
N = 10
x = x_guess
print("x_0 = " + str(x))
for n in range(1, N):
    x = 0.5*(x + r/x)
    print("x_{:d} = ".format(n) + str(x))
print("\n r = " + str(r))
print("(x_{:d})^2 = ".format(n) + str(x^2))
```

We now prove a proposition showing the relationship between boundedness and convergence.

PROPOSITION 2.18. *Suppose that the sequence (z_n) converges to w. Then (z_n) is bounded.*

PROOF. Imagine playing the convergence game starting with a choice of $d = 1$. Since the sequence converges to w, I must be able to win, so I can choose an index N such that if $n \geq N$ then $|z_n - w| < 1$. That is, all terms of the sequence z_n with $n \geq N$ are contained in the disc of radius 1 centered at the point w (see Figure 2.6). Note that this disc is itself contained within the larger disc of radius $|w| + 1$ centered at the origin. Now set $B = \max\{|w| + 1, |z_1|, |z_2|, \ldots, |z_{N-1}|\}$. Then $B > 0$ is a bound for the sequence. Indeed, all points z_n with $n \geq N$ (the ones in the small disc) are bounded by $|w| + 1$, hence by the maximum B. This leaves the finitely many initial terms $z_1, z_2, \ldots, z_{N-1}$, which are also bounded by the maximum B. $\square$

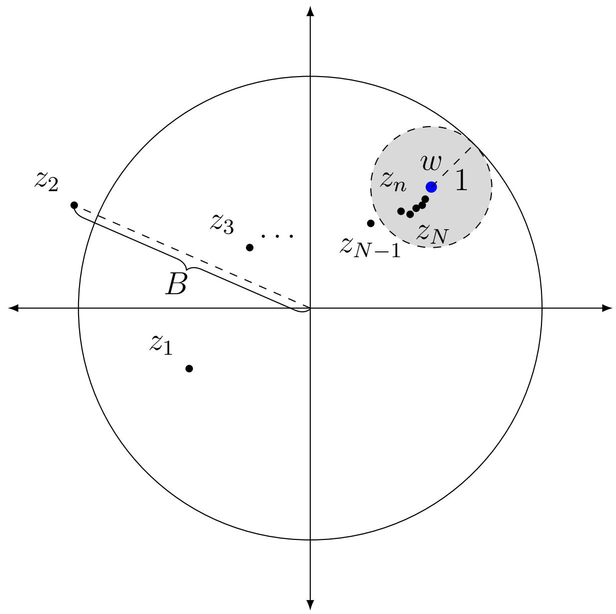


FIGURE 2.6. Convergent implies bounded

REMARK 2.19. Note that there are bounded sequences that do not converge. For example, the sequence $(0, -1, 0, -1, 0, -1, \dots)$ is bounded, but it does not converge: half of the terms are zero and half are -1, so there is no single number w that the terms are getting close to. (See Proposition 2.24 for more about this last assertion.)

To end this section, we briefly review the definition of the Riemann integral of a continuous function from your one-variable calculus course—this provides an excellent example of the central importance of convergent sequences for calculus.

Recall the setup: we have a continuous real-valued function f defined on a closed interval $[a, b]$. For illustrative purposes, let's suppose that f is nonnegative, so that its graph lies entirely above the x-axis (see Figure 2.7). Our motivation is to find the exact area under the graph of f over the interval $[a, b]$. The strategy is a natural one, relying on the fact that we can easily compute the areas of rectangles using the formula *area* = *base* × *height*. We make a sequence of approximations $(R_1, R_2, R_3, \dots)$ to the true area. Here is how we produce the nth approximation R_n, called the *nth Riemann*

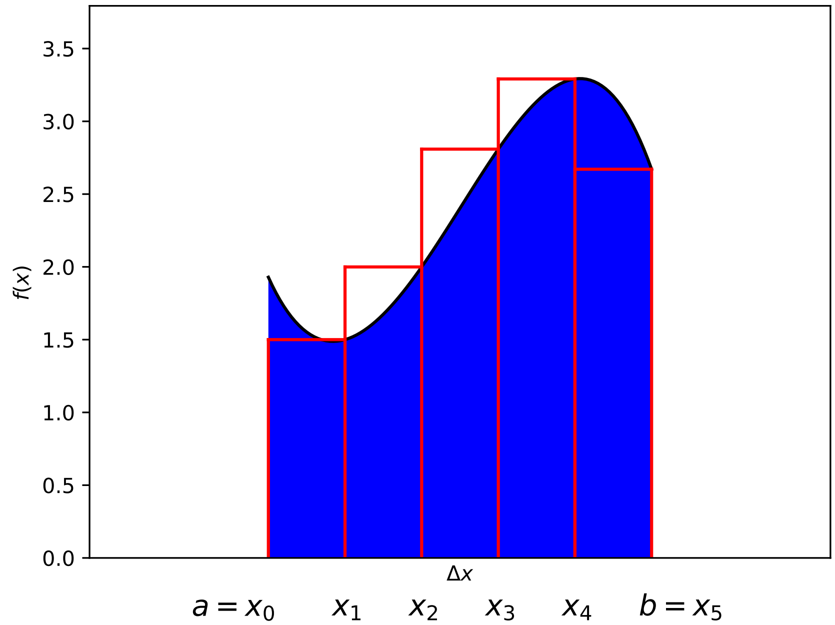


FIGURE 2.7. The Riemann integral computes the blue area under the graph of f over the interval $[a, b]$. The area of the 5 red rectangles is the fifth Riemann sum R_5; each of these rectangles has width $\Delta x = (b - a)/5$.

sum: divide the interval $[a, b]$ into n equal subintervals with endpoints $a = x_0 < x_1 < x_2 < \cdots < x_n = b$; each subinterval has length $\Delta x = (b - a)/n$. We now approximate the region under the curve by a collection of n rectangles. The jth rectangle has base $[x_{j-1}, x_j]$ and height $f(x_j)$, hence area $f(x_j)\Delta x$. The sum R_n of the areas of these n rectangles is then an approximation to the true area under the curve:

$$R_n = f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x = \sum_{j=1}^{n} f(x_j)\Delta x.$$

The continuity of the function f guarantees that the sequence (R_n) converges, and we denote the limit by the familiar symbol:

$$\int_a^b f(x)dx = \lim_{n\to\infty} R_n.$$

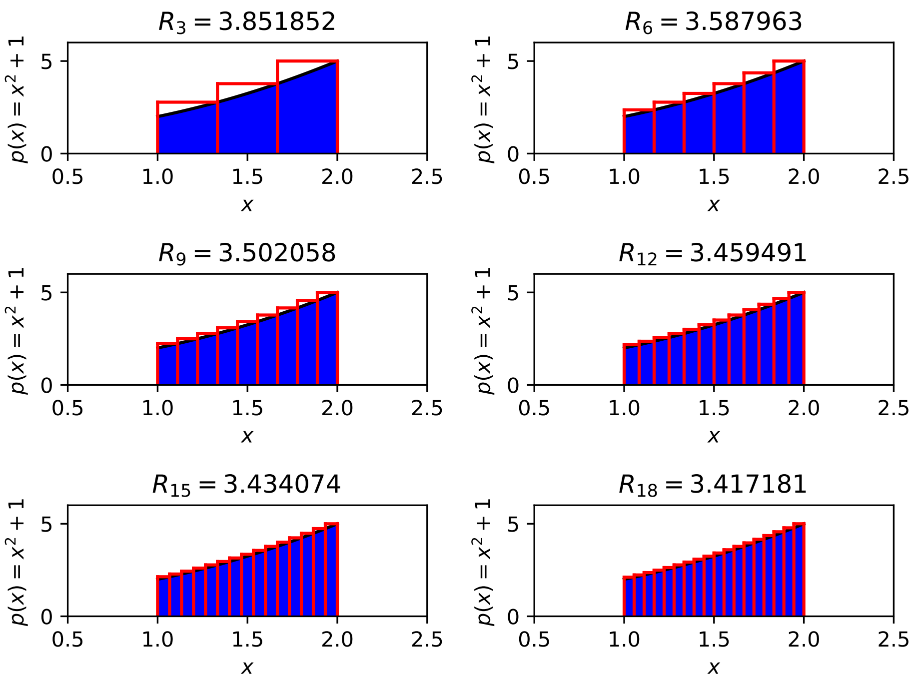


FIGURE 2.8. Riemann sums for the integral of the polynomial $p(x) = x^2 + 1$ over the interval $[1, 2]$.

You will see a proof of this result (the convergence of the sequence) if you take an analysis course, and you will also extend the theory of integration to functions with discontinuities.

Figure 2.8 shows various Riemann sums for the integral of the polynomial $p(x) = x^2+1$ over the interval $[1, 2]$. Here are the values of some later terms in the sequence (to 6 decimal places of precision):

$$\begin{aligned} R_{1000} &\approx 3.334833 \\ R_{1001} &\approx 3.334832 \\ R_{1002} &\approx 3.334831 \\ R_{1003} &\approx 3.334829 \\ R_{1004} &\approx 3.334828 \end{aligned}$$

$$\begin{aligned} R_{1005} &\approx 3.334826 \\ R_{1006} &\approx 3.334825 \\ R_{1007} &\approx 3.334823 \\ R_{1008} &\approx 3.334822 \\ R_{1009} &\approx 3.334820 \end{aligned}$$

The Riemann sums seem to be decreasing, and you might guess that they are converging to $3.333\cdots = 10/3$. Indeed, in this case we can use the Fundamental Theorem of Calculus to compute the integral without having to investigate the limit of the Riemann sums: since $P(x) = \frac{1}{3}x^3 + x$ is an antiderivative of $p(x) = x^2 + 1$, it follows that

$$\int_1^2 (x^2+1)dx = P(2) - P(1) = \frac{8}{3} + 2 - \frac{1}{3} - 1 = \frac{10}{3}.$$

Key points from Section 2.4:

- Convergent sequences (Definition 2.14) and the convergence game (page 57)
- Convergent implies bounded (Proposition 2.18)

2.5. Limit Properties

In this section, we describe some basic properties of limits. These will allow us to work efficiently with various expressions involving limits, and to compute new limits if we know some old ones. Our central result states that limits behave well with respect to the operations of addition, subtraction, multiplication, and division. These properties should remind you of the limit laws for functions from calculus.

Proposition 2.20 (Limit Laws). *Suppose that* $\lim_{n\to\infty} z_n = z$ *and also that* $\lim_{n\to\infty} w_n = w$. *Then*

(a) $\lim_{n\to\infty}(z_n + w_n) = \lim_{n\to\infty} z_n + \lim_{n\to\infty} w_n = z + w$;
(b) $\lim_{n\to\infty}(z_n - w_n) = \lim_{n\to\infty} z_n - \lim_{n\to\infty} w_n = z - w$;
(c) if c *is a fixed complex number, then*

$$\lim_{n\to\infty}(cz_n) = c \lim_{n\to\infty}(z_n) = cz;$$

(d) $\lim_{n\to\infty}(z_n w_n) = (\lim_{n\to\infty} z_n)(\lim_{n\to\infty} w_n) = zw$;
(e) if $w \neq 0$, *then* $\lim_{n\to\infty}(z_n/w_n) = z/w$. *More precisely: there exists an index* N *such that for* $n \geq N$, *the terms* $w_n \neq 0$, *and the sequence* $(z_N/w_N, z_{N+1}/w_{N+1}, z_{N+2}/w_{N+2}, \ldots)$ *converges to* z/w.

We hope these properties are intuitively clear (although we will present the proof of part (a) below). For instance, part (a) roughly says that if the numbers z_n are getting close to z and the numbers w_n are getting close to w, then the sums $z_n + w_n$ are getting close to $z + w$.

EXAMPLE 2.21. Here is a simple but typical use of Proposition 2.20: we will show that the sequence of reciprocal squares

$$\left(\frac{1}{n^2}\right) = \left(1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, \dots\right)$$

converges to zero, starting with the fact (established in Section 2.2) that the harmonic sequence $(1/n)$ converges to zero. Note that the nth term of the sequence, $1/n^2$, is the square of the nth term of the harmonic sequence, so we can apply part (d) of Proposition 2.20 with $z_n = w_n = 1/n$. We find that

$$\lim_{n\to\infty} \frac{1}{n^2} = \lim_{n\to\infty} \left(\frac{1}{n} \cdot \frac{1}{n}\right) = \left(\lim_{n\to\infty} \frac{1}{n}\right) \cdot \left(\lim_{n\to\infty} \frac{1}{n}\right) = 0 \cdot 0 = 0.$$

Repeating this argument shows that for any integer exponent $p \geq 1$, we have $\lim_{n\to\infty} 1/n^p = 0$.

EXAMPLE 2.22. Now we use the limit laws to investigate the sequence $\left(\frac{n-1}{n+1} - \frac{1}{n}\right)$. From Example 2.15, we know that

$$\lim_{n\to\infty} \frac{n-1}{n+1} = 1.$$

We also know that $\lim_{n\to\infty} 1/n = 0$. So we use part (b) with $z_n = \frac{n-1}{n+1}$ and $w_n = \frac{1}{n}$:

$$\lim_{n\to\infty} \left(\frac{n-1}{n+1} - \frac{1}{n}\right) = \lim_{n\to\infty} \frac{n-1}{n+1} - \lim_{n\to\infty} \frac{1}{n} = 1 - 0 = 1.$$

To prove the results of Proposition 2.20 rigorously from the formal definition of convergence requires a bit of technical skill, so we will prove only part (a) to give the flavor of the arguments, and omit the other proofs. The basic idea goes as follows: since $\lim_{n\to\infty} z_n = z$, I can win the convergence game for (z_n) no matter what distance I am challenged with. The same goes for the convergent sequence (w_n). I will use my ability to win these two different convergence games in order to win the convergence game for the sequence of sums $(z_n + w_n)$. Now for the details:

PROOF OF PROPOSITION 2.20, PART (A). As usual, we play the convergence game to show that the sequence $(z_n + w_n)$ converges to $z + w$. You begin by choosing a distance $d > 0$, and I must find an index N such that for all indices $n \geq N$, we have $|z_n + w_n - (z+w)| < d$. Let's begin with some scratch work:

Scratch work: Quantity I want to be smaller than d:

$$|z_n + w_n - (z+w)| = |(z_n - z) + (w_n - w)| \leq |z_n - z| + |w_n - w|,$$

using the triangle inequality (Proposition 1.7). This suggests that we should make sure that the terms $|z_n - z|$ and $|w_n - w|$ are each smaller than $d/2$. I now have an idea for how to choose N, and I can exit the scratchwork.

To choose my index N, I first play the convergence game for (z_n), but using the smaller distance $d/2$. Since (z_n) converges to z, I can choose an index N_1 such that if $n \geq N_1$, then $|z_n - z| < d/2$. Similarly for (w_n), I can choose an index N_2 such that if $n \geq N_2$ then $|w_n - w| < d/2$. I now announce my choice of index $N = \max\{N_1, N_2\}$, which I claim works for the sequence of sums $(z_n + w_n)$.

It is your turn to verify. So suppose that $n \geq N$, which means that $n \geq N_1$ and $n \geq N_2$. Using the triangle inequality (Proposition 1.7), we have

$$\begin{aligned}
|z_n + w_n - (z+w)| &= |(z_n - z) + (w_n - w)| \\
&\leq |z_n - z| + |w_n - w| \\
&< \frac{d}{2} + \frac{d}{2} \\
&= d.
\end{aligned}$$

This concludes your verification of my index N, and since I can find such an N for any d you choose, the sequence $(z_n + w_n)$ converges to $z + w$. □

REMARK 2.23. Be careful to check the hypotheses carefully before applying the limit laws. Here are some common mistakes:

- The sequence of sums $(z_n + w_n)$ may not converge if both of the sequences (z_n) and (w_n) don't converge. For instance: if $(z_n) = (n)$ and $(w_n) = (1/n)$, then (z_n) is unbounded (hence not convergent), and so is $(z_n + w_n) = (n + \frac{1}{n})$.
- However, if neither (z_n) nor (w_n) converge, it is *possible* that $(z_n + w_n)$ *does* converge. For instance, suppose that $(z_n) = (n)$ and $(w_n) = (i - n)$. Then both sequences are unbounded (hence not convergent), but their sum is a constant sequence $(z_n + w_n) = (n + i - n) = (i)_{n\geq 1}$, which converges to i.

EXERCISE 2.7. Suppose that (z_n) and (w_n) are complex sequences, and suppose that (w_n) and $(z_n + w_n)$ are both convergent sequences. Use part (b) of Proposition 2.20 to prove that (z_n) must be convergent as well.

We can use the limit laws to give an easy proof of the fact that limits are unique:

PROPOSITION 2.24 (Uniqueness of Limits). *Suppose that the sequence (z_n) converges to w and to z. That is, suppose that*

$$\lim_{n\to\infty} z_n = z \qquad \text{and} \qquad \lim_{n\to\infty} z_n = w.$$

Then $z = w$.

PROOF. We use part (b) of Proposition 2.20 in reverse order:

$$z - w = \lim_{n\to\infty} z_n - \lim_{n\to\infty} z_n = \lim_{n\to\infty} (z_n - z_n) = \lim_{n\to\infty} (0) = 0.$$

It follows that $z = w$ as claimed. □

As an important direct application of Proposition 2.20, we prove the following result which says that the convergence of a complex sequence is determined by the convergence of its real and imaginary parts.

PROPOSITION 2.25. *Suppose that $(z_n) = (a_n + b_n i)$ is a sequence of complex numbers, with corresponding sequences of real and imaginary* *parts* $(a_n) = (\mathrm{Re}(z_n))$ *and* $(b_n) = (\mathrm{Im}(z_n))$*. Then* $\lim_{n\to\infty} z_n = a + bi$ *if and only if*

$$\lim_{n\to\infty} a_n = a \qquad \textit{and} \qquad \lim_{n\to\infty} b_n = b.$$

Proof. The phrase "if and only if" means that we need to prove both directions: the convergence of (z_n) implies the convergence of (a_n) and (b_n), and vice versa.

First suppose that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. By Proposition 2.20(c), the sequence $(b_n i)$ converges to bi, and then by part (a) we have

$$\lim_{n\to\infty} z_n = \lim_{n\to\infty} (a_n + b_n i) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} (b_n i) = a + bi.$$

For the other direction, suppose that $\lim_{n\to\infty} z_n = a + bi$. We will show that the sequence of real parts (a_n) converges to a, and an analogous argument works for the imaginary parts. We play the convergence game: you challenge me with a small positive number $d > 0$ and I must choose an index N. Well, I know that (z_n) converges to $a + bi$, so I choose N such that if $n \geq N$, then $|z_n - (a + bi)| < d$. But $a_n - a = \mathrm{Re}(z_n - (a + bi))$, so by Proposition 1.6

$$|a_n - a| \leq |z_n - (a + bi)| < d.$$

Since this is true for all $n \geq N$, it follows that my index is valid, and this means that the sequence (a_n) converges to a as claimed. □

Remark 2.26. The previous result implies that, in principle, we can investigate the convergence of complex sequences (z_n) by instead investigating the convergence of the real sequences $(\mathrm{Re}(z_n))$ and $(\mathrm{Im}(z_n))$. However, this is generally not the way we will prove the convergence of complex sequences. Nevertheless, real sequences are important in their own right, and we will sometimes develop results and techniques that apply only to real sequences. The point is not that we have lost interest in complex sequences, but rather that some important results hold only for real sequences and will ultimately help us in our study of more general complex sequences.

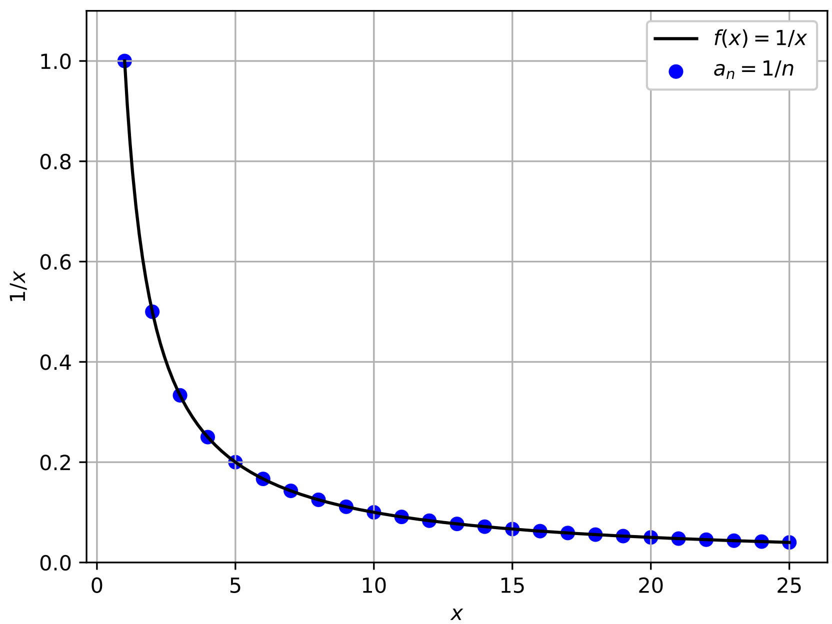


FIGURE 2.9. Graph of $f(x) = 1/x$ and the harmonic sequence

If $f(x)$ is any real function defined on the positive real numbers, then we can manufacture an associated sequence $(a_n) = (f(n))$ by evaluating at the positive integers.

EXAMPLE 2.27. Consider the reciprocal function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = 1/x$. Then setting $a_n = f(n)$ yields the harmonic sequence $(1/n)$ (see Figure 2.9).

EXAMPLE 2.28. Now consider the arctangent function

$$\arctan : \mathbb{R} \to (-\pi/2, \pi/2).$$

Setting $a_n = \arctan(n)$ yields the sequence $(\arctan(n))$ (see Figure 2.10).

Whenever a real sequence (a_n) arises from a real function $f(x)$ in this way, there is a close relationship between the limit of the sequence and the limit of the function at infinity. Before stating this precisely, we need to formally introduce the notion of an infinite limit for a real sequence; this is a particular type of divergence.

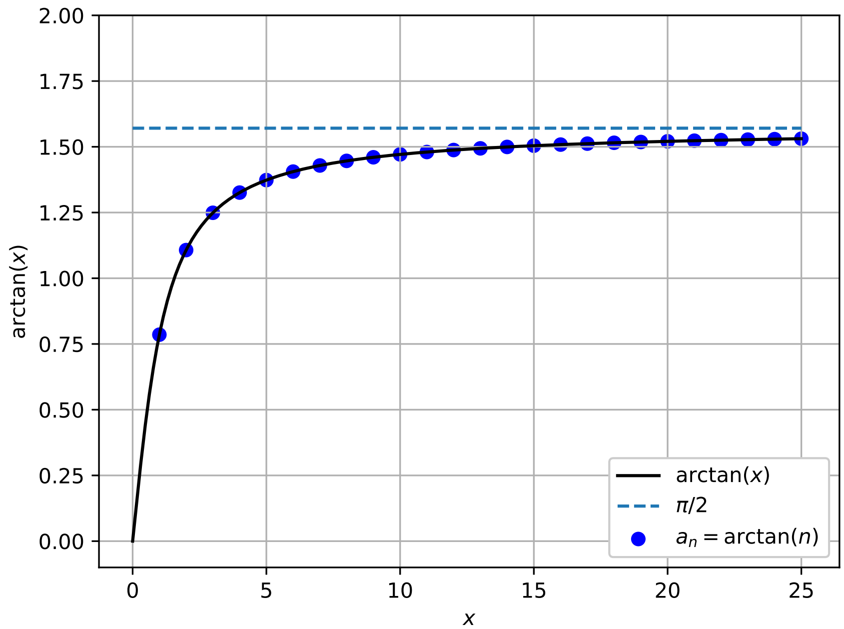


FIGURE 2.10. Graph of arctan(x) and the arctangent sequence

DEFINITION 2.29. Suppose that (a_n) is a sequence of real numbers. We write $\lim_{n\to\infty} a_n = +\infty$ if for all positive numbers $B > 0$, there exists an index $N > 0$ such that $n \geq N$ implies that $a_n > B$. In this case, we say that the sequence *diverges to* $+\infty$.

Similarly, we write $\lim_{n\to\infty} a_n = -\infty$ if for all negative numbers $B < 0$, there exists an index $N > 0$ such that $n \geq N$ implies that $a_n < B$. In this case, we say that the sequence *diverges to* $-\infty$.

EXERCISE 2.8. What is the difference between saying that the real sequence (a_n) is unbounded and saying that $\lim_{n\to\infty} a_n = \pm\infty$? Can you give an example of a positive sequence (a_n) that is unbounded but which does not diverge to $+\infty$?

PROPOSITION 2.30. *Suppose that $f(x)$ is a real-valued function defined on the positive x-axis, and consider the associated sequence*

$(a_n) = (f(n))$. *Suppose that* $\lim_{x\to\infty} f(x) = L$, *where* L *is either a real number or* $\pm\infty$. *Then* $\lim_{n\to\infty} a_n = L$.

PROOF. We treat the case where L is a real number, the cases $\pm\infty$ being similar. In your first calculus course, you may have studied function limits in a less formal way than we are studying limits of sequences. In that course, you may have described function limits at infinity as follows: to say that the real number L is the limit of $f(x)$ as $x \to \infty$ means that the values $f(x)$ get arbitrarily close to L as x gets large. Restricting attention to integral values $x = n$, we find that the terms $a_n = f(n)$ get arbitrarily close to L as n gets large. This is exactly the informal definition of sequence convergence! EXERCISE 2.9 below challenges you to make this argument rigorous. □

REMARK 2.31. The converse of the previous proposition is false in general. That is, the behavior of the sequence (a_n) need not determine the behavior of the function $f(x)$. For instance, consider the cosine function $f(x) = \cos(2\pi x)$. Then the associated sequence is constant: $(a_n) = (\cos(2\pi n)) = (1)_{n\geq 1}$. Thus, we have $\lim_{n\to\infty} a_n = 1$. But the function $f(x)$ continues to oscillate between the extremes of ± 1 as $x \to \infty$, so it does not converge to a limit or diverge to $\pm\infty$.

EXERCISE 2.9. Let $f\colon (0,\infty) \to \mathbb{R}$ be a real function and L a real number. In your calculus course, you may have seen the following formal definition for the statement that $\lim_{x\to\infty} f(x) = L$:

For every positive real number $d > 0$, *there exists a positive real number* $M > 0$ *such that for all* $x \geq M$, *we have* $|f(x) - L| < d$.

(a) Write a short paragraph explaining how this formal definition corresponds to a more intuitive understanding of the limiting behavior depicted in Figures 2.9 and 2.10:

$$\lim_{x\to\infty} \frac{1}{x} = 0 \qquad \text{and} \qquad \lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}.$$

(b) Use the formal definition provided above together with Definition 2.14 to give a rigorous proof of Proposition 2.30 in the case where L is a real number.

We can use the connection between sequence limits and function limits at infinity to find the limits of many real sequences.

EXAMPLE 2.32. Look back at Figure 2.10 showing the arctangent function and the associated sequence. The graph has a horizontal asymptote at $\pi/2$, which means that $\lim_{x\to\infty} \arctan(x) = \pi/2$. It follows that $\lim_{n\to\infty} \arctan(n) = \pi/2$ as well.

EXAMPLE 2.33. Consider the sequence $(\ln(n)/\sqrt{n})$ arising from the function $f(x) = \ln(x)/\sqrt{x}$. Since both $\lim_{x\to\infty} \ln(x) = +\infty$ and $\lim_{x\to\infty} \sqrt{x} = +\infty$, we will use L'Hôpital's rule to compute the limit of f. So begin by investigating the ratio of derivatives:

$$\frac{(\ln(x))'}{(\sqrt{x})'} = \frac{1/x}{(1/2\sqrt{x})} = \frac{2\sqrt{x}}{x} = \frac{2}{\sqrt{x}} \longrightarrow 0 \quad \text{as } x \to \infty.$$

L'Hôpital implies that $\lim_{x\to\infty} f(x) = 0$, so $\lim_{n\to\infty}(\ln(n)/\sqrt{n}) = 0$.

EXAMPLE 2.34 (p-sequences). Fix a real exponent p and consider the sequence $(a_n) = (1/n^p)$. Then

$$\lim_{n\to\infty} \frac{1}{n^p} = \begin{cases} 0 & p > 0 \\ 1 & p = 0 \\ +\infty & p < 0. \end{cases}$$

This follows from the limiting behavior of the functions $f(x) = 1/x^p$ as $x \to \infty$ (see Figure 2.11).

EXAMPLE 2.35 (Real Geometric Sequences). Consider the geometric sequence $(a_n) = \left(\frac{1}{2^n}\right)$ associated to the function $f(x) = 1/2^x$. Since $\lim_{x\to\infty} \frac{1}{2^x} = 0$, it follows that $\lim_{n\to\infty} \frac{1}{2^n} = 0$. More generally, we can use this approach to determine the behavior of all real geometric sequences.

Fix a nonnegative real number $r \geq 0$, and consider the geometric sequence $(a_n) = (r^n)$. Then

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & 0 \leq r < 1 \\ 1 & r = 1 \\ +\infty & r > 1. \end{cases}$$

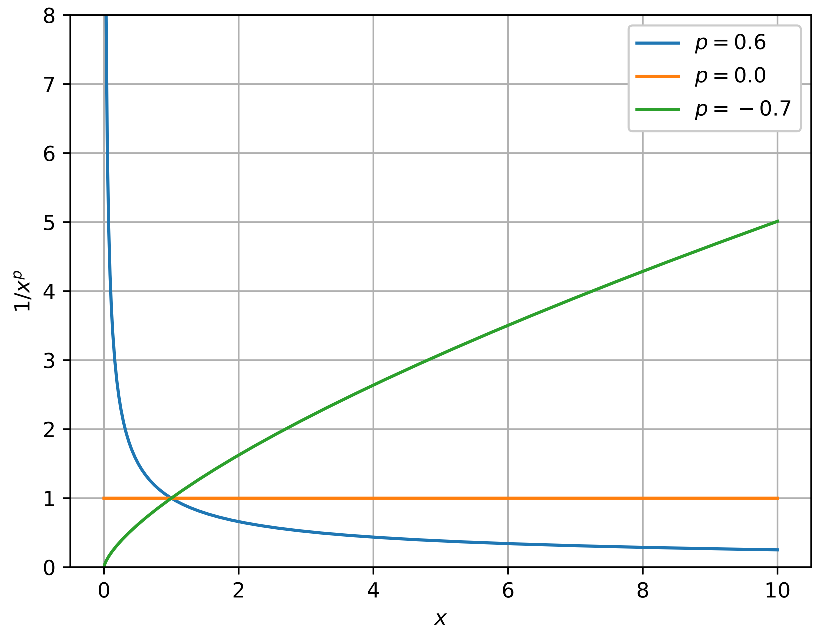


FIGURE 2.11. Graphs of $1/x^p$ for various p.

This follows from the limiting behavior of the functions $f(x) = r^x$ as $x \to \infty$. (see Figure 2.12).

As an example of using real sequences to learn about complex sequences (Remark 2.26), we have the following simple but useful result.

PROPOSITION 2.36. *A sequence of complex numbers* (z_n) *converges to zero if and only if the real sequence of magnitudes* $(|z_n|)$ *converges to zero.*

PROOF. Here is an informal proof: to say that the sequence of complex numbers (z_n) converges to zero is to say that the distances of the numbers z_n from zero get arbitrarily small. But these distances are exactly the magnitudes $|z_n|$, so saying that the magnitudes converge to zero is equivalent to saying that the numbers themselves converge to zero. The next exercise asks you to make this argument rigorous, using the formal definition of convergence. $\square$

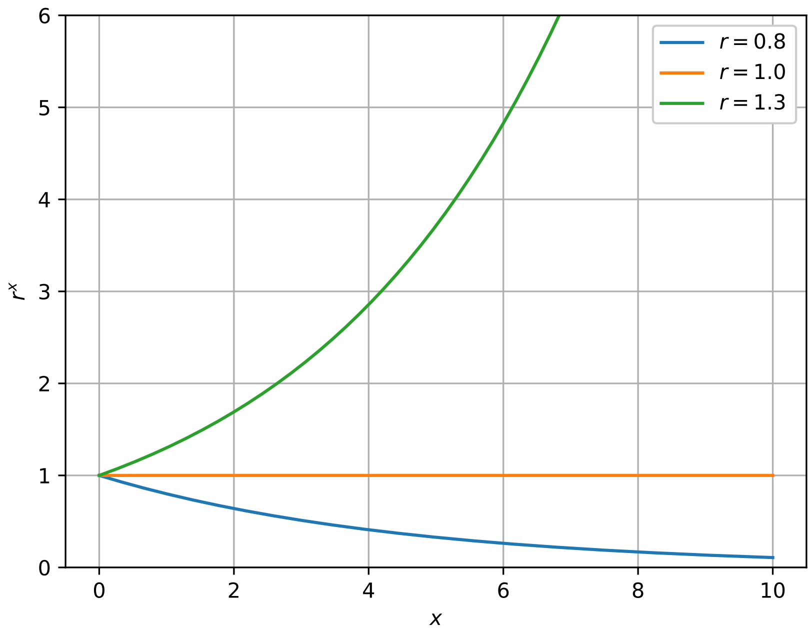


FIGURE 2.12. Graphs of r^x for various r.

EXERCISE 2.10. Use Definition 2.14 to give a careful proof of Proposition 2.36. The key point to notice is that $||z_n| - 0| = |z_n - 0|$, so playing the convergence game for $(|z_n|)$ is the same as playing the convergence game for (z_n). Be sure to prove both directions: if $\lim_{n\to\infty} z_n = 0$ then $\lim_{n\to\infty} |z_n| = 0$, and vice-versa.

EXAMPLE 2.37 (Alternating p-sequences). Fix a real exponent p and consider the *alternating p-sequence*

$$\left(\frac{(-1)^{n+1}}{n^p} \right) = \left(1, -\frac{1}{2^p}, \frac{1}{3^p}, -\frac{1}{4^p}, \dots \right).$$

First suppose $p > 0$. From Example 2.34, we know that the corresponding sequence of magnitudes $(1/n^p)$ converges to zero, so it follows from Proposition 2.36 that

$$\lim_{n\to\infty} \frac{(-1)^{n+1}}{n^p} = 0 \qquad \text{if } p > 0.$$

If $p < 0$, then the sequence of magnitudes $(1/n^p)$ diverges to $+\infty$, and so the alternating sequence is unbounded, hence does not converge.

For $p = 0$ we have the divergent sequence $(1, -1, 1, -1, \dots)$.

EXAMPLE 2.38 (Complex Geometric Sequences). Consider the geometric sequence $(c_n) = \left(\left(\frac{i}{2}\right)^n\right)$. The corresponding sequence of magnitudes is $(|c_n|) = \left(\frac{1}{2^n}\right)$ which converges to zero by Example 2.35. It follows from Proposition 2.36 that $\lim_{n\to\infty} \left(\frac{i}{2}\right)^n = 0$ as well. As before, we can use this approach to determine the behavior of all complex geometric sequences.

Fix a complex number c and consider the geometric sequence (c^n). First suppose $|c| < 1$. From Example 2.35, the corresponding sequence of magnitudes $(|c^n|) = (|c|^n)$ converges to zero, so Proposition 2.36 says that

$$\lim_{n\to\infty} c^n = 0 \qquad \text{if } |c| < 1.$$

If $|c| > 1$, then the sequence of magnitudes $(|c|^n)$ diverges to $+\infty$, and so the original sequence (c^n) is unbounded, hence divergent.

For $|c| = 1$, the complex number c lies on the unit circle. You investigated these sequences in EXERCISE 2.2. Except for the case $c = 1$, none of these sequences converge, but instead rotate around the unit circle forever.

Key points from Section 2.5:

- Limit laws (Proposition 2.20)
- Real and imaginary parts of complex limits (Proposition 2.25)
- Relation between function limits at infinity and sequence limits (Proposition 2.30)
- Behavior of p-sequences (Examples 2.34 and 2.37) and geometric sequences (Examples 2.35 and 2.38)

2.6. The Monotone Convergence Theorem ($\mathbb{R}$)

This section takes place entirely in the context of the real numbers $\mathbb{R}$.

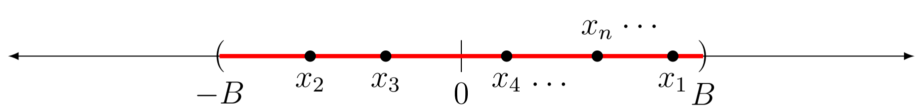
As mentioned in Remark 2.26 from the previous section, we will often develop strong results that apply only to real sequences, and then later use those results to study more general complex sequences. This section is devoted to an important result of this type, called the *Monotone Convergence Theorem*. It applies only to a special kind of real sequence, presented in the following definition.

DEFINITION 2.39. A sequence of real numbers (x_n) is *increasing* if $x_n \leq x_{n+1}$ for all indices $n \geq 1$. Similarly, the sequence is *decreasing* if $x_n \geq x_{n+1}$ for all indices $n \geq 1$. The sequence is *monotone* if it is either increasing or decreasing.

EXAMPLE 2.40. Here are two examples and a non-example:

(a) The sequence of counting numbers $(n) = (1, 2, 3, \dots)$ is increasing;
(b) The harmonic sequence $(1/n) = (1, 1/2, 1/3, \dots)$ is decreasing;
(c) The alternating harmonic sequence $(1, -1/2, 1/3, -1/4, \dots)$ is not monotone.

REMARK 2.41. Recall that a general complex sequence (z_n) is bounded if it is entirely contained within a finite disc (see Figure 2.3). Specializing to real sequences (x_n), we see that (x_n) is bounded if it is entirely contained within a finite interval $[-B, B]$:



In the context of real sequences, two additional notions of boundedness are often useful: we say that (x_n) is *bounded above* if there is an *upper bound* U such that $x_n \leq U$ for all n. Likewise, we way that (x_n) is *bounded below* if there is an *lower bound* L such that $L \leq x_n$ for all n. Here are three simple examples:

- the sequence of counting numbers $(n) = (1, 2, 3, \dots)$ is bounded below (by $L = 0$ for example) but not bounded above;

- the sequence of negative counting numbers defined by $(-n) = (-1, -2, -3, \dots)$ is bounded above (by $U = -1$ for example) but not bounded below.
- the harmonic sequence $(1/n)$ is bounded above by $U = 1$ and bounded below by $L = 0$.

EXERCISE 2.11. Show that a real sequence is bounded exactly when it is bounded above and bounded below.

EXERCISE 2.12. Suppose that U is an upper bound for a convergent real sequence (a_n), so that $a_n \leq U$ for all n. Show that $\lim_{n\to\infty} a_n \leq U$.

This new language is especially useful for monotone sequences (x_n):

- If (x_n) is increasing, then the first term x_1 provides a lower bound for the sequence. Hence to prove that an increasing sequence is bounded, we just need to find an upper bound.
- Similarly, if (x_n) is decreasing, then the first term x_1 provides an upper bound for the sequence. So to prove that a decreasing sequence is bounded, we just need to find a lower bound.

EXAMPLE 2.42. It is often not immediately obvious whether a sequence is monotone. For instance, consider the sequence (x_n) with

$$x_n = \frac{2n-3}{2n+3}.$$

Here is the beginning of the sequence, starting with the index $n = 1$:

$$(-1/5,\ 1/7,\ 3/9,\ 5/11,\ 7/13,\ 9/15,\ \dots).$$

These terms are increasing, but are we sure that this continues? We need to prove that for each integer $n \geq 1$ the inequality $x_n \leq x_{n+1}$ is valid; so far we have only checked the first few cases. Instead of proving $x_n \leq x_{n+1}$, we will prove the equivalent statement $x_{n+1} - x_n \geq 0$. The proof consists of carefully manipulating the general expression: for any value of $n \geq 1$, we have

$$\begin{aligned} x_{n+1} - x_n &= \frac{2(n+1)-3}{2(n+1)+3} - \frac{2n-3}{2n+3} \\ &= \frac{2n-1}{2n+5} - \frac{2n-3}{2n+3} \\ &= \frac{(2n-1)(2n+3)-(2n-3)(2n+5)}{(2n+5)(2n+3)} \\ &= \frac{4n-3-(4n-15)}{(2n+5)(2n+3)} \\ &= \frac{12}{(2n+5)(2n+3)} \\ &> 0. \end{aligned}$$

This shows that the sequence (x_n) is increasing.

Note that the sequence is also bounded above (hence bounded): the numerator $2n-3$ is always less than the denominator $2n+3$, so $x_n < 1$ for all n.

Finally, let's use the limit laws to show that (x_n) converges; be sure you can justify each step using Proposition 2.20:

$$\begin{aligned} \lim_{n\to\infty} \frac{2n-3}{2n+3} &= \lim_{n\to\infty} \frac{2-\frac{3}{n}}{2+\frac{3}{n}} \\ &= \frac{\lim_{n\to\infty}(2-\frac{3}{n})}{\lim_{n\to\infty}(2+\frac{3}{n})} \\ &= \frac{2-3\lim_{n\to\infty}\frac{1}{n}}{2+3\lim_{n\to\infty}\frac{1}{n}} \\ &= \frac{2-3\cdot 0}{2+3\cdot 0} \\ &= 1. \end{aligned}$$

In the previous example we looked at a real sequence (x_n) having all three of the following properties:

- monotone;
- bounded;
- convergent.

We already know about the relationship between boundedness and convergence: all convergent sequences are bounded (Proposition 2.18) but not all bounded sequences converge. What about the relationship between monotonicity and convergence? Well, there is no immediate implication in either direction:

- The alternating harmonic sequence $((-1)^{n+1}/n)$ converges to zero but is not monotone;
- The sequence $(n) = (1, 2, 3, \dots)$ is monotone but does not converge.

However, boundedness plus monotonicity is sufficient to guarantee convergence.

THEOREM 2.43 (Monotone Convergence Theorem). *Suppose that the real sequence (x_n) is monotone and bounded. Then (x_n) converges.*

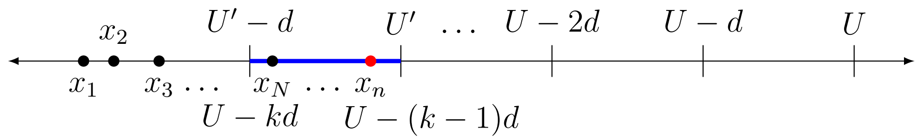
The Monotone Convergence Theorem (MCT) relies on a special property of the real numbers called *completeness*—see the optional Section 2.7 to learn more about completeness and to see a full proof of the MCT. For now, we will provide only a partial argument for the MCT, but one which provides good intuition.

Partial Proof of MCT: Imagine that we have a bounded and increasing real sequence $x_1 \le x_2 \le x_3 \le \dots$. In particular, there exists an upper bound $U > 0$ such that $x_n \le U$ for all n. Think of U as a ceiling: the sequence is going up (or possibly staying the same), but it never gets higher than U. You may already be feeling that the sequence must converge, but just to sharpen that intuition, let's investigate a little further by playing the convergence game.

We don't yet have a candidate limit, but you begin anyway by challenging me with a small positive number $d > 0$. I must respond with an index N, but first I do some preliminary work: consider the decreasing sequence

$$(U, U - d, U - 2d, U - 3d, \dots).$$

Think of this as lowering the ceiling in steps of size d:



There will be a first positive integer k such that $U - kd$ is *not* an upper bound for the sequence (x_n). Having found this integer k, set

$$U' = U - (k-1)d$$

so that U' is an upper bound for the sequence, but $U' - d$ is not an upper bound. Having adjusted the ceiling in this manner, I am now ready to choose my index N.

Since $U' - d$ is not an upper bound, there exists an index N such that $x_N > U' - d$; this is my chosen index. So consider any index $n \geq N$. The sequence is increasing, so we have $x_N \leq x_n$. But U' is an upper bound for the sequence, so putting all of this together yields

$$U' - d < x_N \leq x_n \leq U'.$$

This means that all but the first $N - 1$ terms of the sequence are contained in the half-open interval $(U' - d, U']$, which has length d.

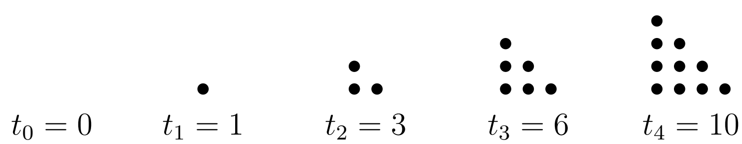
The upshot: no matter how small of a distance $d > 0$ you specify, all but finitely many terms of the sequence are huddled together in an interval of size d. Since they are all so close to each other, it is plausible that they are getting close to some particular real number a, which would then be their limit. The existence of the limit a is a consequence of the *completeness* of the real numbers, as explained in the optional Section 2.7.

The MCT is a powerful tool for proving the convergence of monotone sequences (x_n): instead of playing the convergence game, we simply need to show that (x_n) is monotone and bounded. In Example 2.42, we saw how to establish these properties by working directly with an explicit formula for the general term x_n. But many sequences have a recursive definition, and we may not have an explicit formula to work with. In these situations, we will often use a proof technique called *induction.* We introduce this technique with the following example, and then we will discuss it more formally.

EXAMPLE 2.44. Consider the recursively defined sequence (t_n) defined by

$$t_0 = 0 \qquad \text{and} \qquad t_n = t_{n-1} + n \qquad \text{for } n \geq 1.$$

The following picture indicates why these are called "triangular numbers."



We are going to show that for all $n \geq 0$, we have $t_n = \frac{n(n+1)}{2}$. Our proof will be an example of the induction technique.

Here is the idea: we want to prove that for all $n \geq 0$, the following equality is true:

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

Note that this is really an infinite list of equalities, one for each value of n. We begin by checking that the *base case* $n = 0$ is valid:

$$t_0 = 0 = \frac{0 \cdot (0+1)}{2}.$$

So now we know the case $n = 0$, and we want to prove the case $n = 1$. And after that, we will want to prove the case $n = 2$, etc. Induction allows us to make a single argument that accounts for all of these at once: we assume that the equality holds for some generic value $n = k$, and then we use that information to prove that the inequality must also hold for the next value $n = k + 1$.

So suppose that $t_k = \frac{k(k+1)}{2}$ for some value $n = k$. This assumption is called the *induction hypothesis.* Then use the recursive definition to find t_{k+1}:

$$\begin{aligned} t_{k+1} &= t_k + (k+1) \\ &= \frac{k(k+1)}{2} + (k+1) \\ &= \frac{k(k+1)+2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned}$$

This is the desired equality for $n = k+1$. The *principle of mathematical induction* below now says that the stated equalities hold for all $n \geq 0$.

The previous example introduced the technique of induction in order to establish an infinite list of equalities. This technique rests on the following fact about the natural numbers.

Principle of Mathematical Induction: Suppose that $(\mathcal{P}(n))$ is a sequence of mathematical statements, one for each natural number $n \geq 0$. Moreover, suppose that

- The statement $\mathcal{P}(0)$ is true;
- for every $n \geq 0$, the truth of statement $\mathcal{P}(n)$ implies the truth of statement $\mathcal{P}(n+1)$.

Then $\mathcal{P}(n)$ is true for every $n \geq 0$.

In the previous example, $\mathcal{P}(n)$ was the equality $t_n = \frac{n(n+1)}{2}$. The proof began by establishing the truth of the base case $\mathcal{P}(0)$. Then we made the *induction step* by proving that if the equality $\mathcal{P}(n)$ is true for some value n, then the equality $\mathcal{P}(n+1)$ is also true.

As a second example of induction, we prove the factorial formula for the binomial coefficients announced on page 47. Recall that the integer $\binom{n}{k}$ is the coefficient of z^k in the expansion of the polynomial $(1+z)^n$. Moreover, in Proposition 2.10, we established the Pascal triangle recursion

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \qquad \text{for } 0 < k < n.$$

PROPOSITION 2.45. *For $0 \leq k \leq n$, we have*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

PROOF. There are two indices here, n and k, so we must decide which one to use for induction. We will use the index n, and here is the statement $\mathcal{P}(n)$ that we wish to prove for every $n \geq 0$:

$$\mathcal{P}(n): \qquad \text{for } 0 \leq k \leq n, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Let's begin by checking explicitly that $\mathcal{P}(0)$ is true: if $n = 0$, then the only value for k is $k = 0$, and we have

$$\frac{0!}{0!(0-0)!} = \frac{1}{1 \cdot 1} = 1 = \binom{0}{0}.$$

Now we assume the induction hypothesis that $\mathcal{P}(n)$ is true for some value of $n \geq 0$, which means that we are assuming the factorial formula holds for all values $0 \leq k \leq n$. We now need to prove that $\mathcal{P}(n+1)$ is true. We start by explicitly checking the two cases $k = 0$ and $k = n+1$:

$$k = 0: \quad \frac{(n+1)!}{0!(n+1-0)!} = \frac{(n+1)!}{1 \cdot (n+1)!} = 1 = \binom{n+1}{0}$$

and

$$k = n+1: \quad \frac{(n+1)!}{(n+1)!(n+1-(n+1))!} = \frac{(n+1)!}{(n+1)! \cdot 0!} = 1 = \binom{n+1}{n+1}.$$

So far so good. Now consider any intermediate value for k, where $0 < k < n+1$. We make use of the recursion from Proposition 2.10:

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

By the induction hypothesis, the factorial formula holds for the binomial coefficients on the right hand side. We write this out explicitly, and then do some careful algebra with factorials. The thing to remember when following this computation is that $m! = m \cdot (m-1)!$.

$$\begin{aligned}
\binom{n+1}{k} &= \binom{n}{k-1} + \binom{n}{k} \\
&= \frac{n!}{(k-1)!(n-(k-1))!} + \frac{n!}{k!(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k+1)!} + \frac{n!}{k!(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{1}{n-k+1} + \frac{1}{k}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{k+(n-k+1)}{k(n-k+1)}\right) \\
&= \frac{n!}{(k-1)!(n-k)!}\left(\frac{n+1}{k(n+1-k)}\right) \\
&= \frac{(n+1)!}{k!(n+1-k)!}.
\end{aligned}$$

Thus, we have shown that the truth of $\mathcal{P}(n)$ implies the truth of $\mathcal{P}(n+1)$, and so by the principle of mathematical induction, the factorial formula holds for all binomial coefficients. □

Now we return to the Monotone Convergence Theorem. In the following example, we use induction to prove that the given sequence is monotone and bounded.

EXAMPLE 2.46. Consider the recursively defined sequence (x_n) with

$$x_0 = 1 \qquad \text{and} \qquad x_n = \frac{1}{3 - x_{n-1}} \qquad \text{for } n \geq 1.$$

Here are the first few terms:

$$(x_n) = \left(1, \frac{1}{2}, \frac{2}{5}, \frac{5}{13}, \frac{13}{34}, \dots\right).$$

These initial terms are in the interval $(0, 1]$ and are decreasing; we want to prove that this continues. We will use induction. We first show that the sequence is bounded. So for each $n \geq 0$, let $\mathcal{P}(n)$ be the following chain of inequalities:

$$0 < x_n \leq 1.$$

The base case $n = 0$ is clearly true, since $x_0 = 1$. For the induction step, we now assume that $0 < x_k \leq 1$ for some value $n = k$. Then multiplying by -1 changes the directions of the inequalities:

$$-1 \leq -x_k < 0.$$

Now add 3 and take reciprocals, which changes the directions again:

$$\frac{1}{3} < \frac{1}{3 - x_k} \leq \frac{1}{2}.$$

The inner expression is the next term x_{k+1}, so we conclude that the statement $\mathcal{P}(k+1)$ is true, namely that $0 < x_{k+1} \leq 1$. By induction, the sequence (x_n) is bounded.

Now we use induction to prove that (x_n) is a decreasing sequence. This time, the statement $\mathcal{P}(n)$ is the inequality $x_{n+1} \leq x_n$. The base case $n = 0$ certainly holds:

$$x_0 = 1 \quad \text{and} \quad x_1 = \frac{1}{2}, \quad \text{so} \quad x_1 < x_0.$$

For the induction step, we now assume that $x_{k+1} \leq x_k$ for some value $n = k$. The proof is similar to the one for boundedness: multiplying by -1 changes the direction of the inequality:

$$-x_{k+1} \geq -x_k.$$

Now add 3:

$$3 - x_{k+1} \geq 3 - x_k.$$

Note that neither side is zero, since we know that all terms of the sequence are between 0 and 1. So we can take reciprocals, which changes the directions again:

$$\frac{1}{3 - x_{k+1}} \leq \frac{1}{3 - x_k}.$$

We recognize these expressions from the recursive definition of the sequence:

$$x_{k+2} \leq x_{k+1}.$$

This is the inequality $\mathcal{P}(k+1)$, and the principle of mathematical induction says that the sequence (x_n) is decreasing. Since (x_n) is also bounded, it converges by the MCT. We set $x = \lim_{n\to\infty} x_n$.

Now that we know the limit x exists, we can use the limit laws together with the recursive definition to find it. First note that we have two slightly different sequences to consider when taking the limit of the recursive definition $x_n = 1/(3 - x_{n-1})$ for $n \geq 1$:

$$(x_n) = (x_1, x_2, x_3, \dots)$$

and

$$(x_{n-1}) = (x_0, x_1, x_2, \dots).$$

The second sequence (x_{n-1}) is the *same* list of numbers in the *same* order, just with the additional term x_0 at the beginning. In any case, these two lists of numbers both converge to x. Taking the limit of the recursive definition yields

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \left(\frac{1}{3 - x_{n-1}} \right) = \frac{1}{3 - \lim_{n\to\infty} x_{n-1}} = \frac{1}{3 - x}.$$

Rearranging, we find that x must be a root of the quadratic polynomial $p(x) = x(x - 3) + 1 = x^2 - 3x + 1$. Using the quadratic formula, we find the two roots of $p(x)$:

$$r_1 = \frac{3 + \sqrt{5}}{2}, \qquad r_2 = \frac{3 - \sqrt{5}}{2}.$$

Note that $r_1 > 2$, while all terms $x_n \leq 1$. So r_1 can't be the limit, and it must be that $x = r_2 = \frac{3-\sqrt{5}}{2}$.

We will need the following result for our next example:

PROPOSITION 2.47 (Inequality of Arithmetic and Geometric Means). *Suppose that $a, b > 0$ are positive real numbers. Then*

$$\sqrt{ab} \leq \frac{1}{2}(a + b).$$

PROOF. Begin by observing that

$$0 \leq (a - b)^2 = a^2 - 2ab + b^2.$$

Now add $4ab$ to both sides and factor:

$$4ab \leq a^2 + 2ab + b^2 = (a + b)^2.$$

Finally, divide by 4 and take the square root:

$$\sqrt{ab} \leq \frac{1}{2}(a+b).$$

□

Example 2.48. Recall the Babylonian sequence from Example 2.17:

$$x_0 = 1 \qquad \text{and} \qquad x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) \quad \text{for } n \geq 1.$$

We have seen numerical evidence that (x_n) converges to $\sqrt{2}$. We now prove this by using the MCT. As a first step, we use the inequality of arithmetic and geometric means to show that for all $n \geq 1$ we have

$$x_n = \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) \geq \sqrt{x_{n-1} \cdot \frac{2}{x_{n-1}}} = \sqrt{2}.$$

Thus, the sequence (x_n) is bounded below by $\sqrt{2}$. (Note that we are ignoring the first term $x_0 = 1$, which does not change the convergence behavior of the sequence.)

Now we will use induction to prove that the sequence is decreasing. We begin with the base case $n = 1$:

$$x_1 = \frac{3}{2} \quad \text{and} \quad x_2 = \frac{1}{2}\left(\frac{3}{2} + \frac{2 \cdot 2}{3}\right) = \frac{3}{4} + \frac{2}{3} < \frac{3}{2} = x_1.$$

For the induction step, assume that $x_k \geq x_{k+1}$ for some value $n = k$. We need to prove that $x_{k+1} \geq x_{k+2}$. For this, we will use some ideas from calculus. Consider the function $f(x) = \frac{1}{2}\left(x + \frac{2}{x}\right)$, so $x_{k+1} = f(x_k)$ and $x_{k+2} = f(x_{k+1})$. The derivative of f is given by

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

This shows that $f'(x) \geq 0$ for all $x \geq \sqrt{2}$. This means that the function is weakly increasing to the right of $\sqrt{2}$: if $\sqrt{2} \leq a \leq b$, then $f(a) \leq f(b)$. But by the induction hypothesis, we have $x_{k+1} \leq x_k$, so we find that $x_{k+2} = f(x_{k+1}) \leq f(x_k) = x_{k+1}$. By the principle of mathematical induction, the sequence is decreasing. Hence, by the MCT, the sequence (x_n) converges to some limit x.

To prove that $x = \sqrt{2}$, we take the limit of the recursion:

$$\begin{aligned} x &= \lim_{n\to\infty} x_n \\ &= \lim_{n\to\infty} \frac{1}{2}\left(x_{n-1} + \frac{2}{x_{n-1}}\right) \\ &= \frac{1}{2}\left(\lim_{n\to\infty} x_{n-1} + \frac{2}{\lim_{n\to\infty} x_{n-1}}\right) \\ &= \frac{1}{2}\left(x + \frac{2}{x}\right). \end{aligned}$$

Rearranging then shows that $x^2 = 2$, so that $x = \sqrt{2}$ as expected.

Remark 2.49. We can use the Monotone Convergence Theorem to provide a justification for what is likely the way you have always thought about real numbers. Namely, if I had asked you 3 weeks ago to tell me what a real number is, you probably would have said something about a decimal expansion, with finitely many digits to the left of the decimal point, but possibly infinitely many to the right:

$$c_1c_2 \cdots c_k.d_1d_2d_3 \cdots$$

For instance, you might have said that π is the decimal expansion with $c_1 = 3$ and $d_1 = 1, d_2 = 4, d_3 = 1, d_4 = 5, d_5 = 9, d_6 = 2, d_7 = 6, \ldots$:

$$\pi = 3.1415926 \cdots$$

Of course, you can never finish telling me the decimal expansion of a real number in this way, but the idea is that every sequence of digits is allowed. The act of telling me a particular decimal expansion is the same as providing a sequence of rational numbers $(x_0, x_1, x_2, \ldots)$:

$$\begin{aligned} x_0 &= c_1 \cdots c_k \\ x_1 &= c_1 \cdots c_k.d_1 \\ x_2 &= c_1 \cdots c_k.d_1d_2 \\ x_3 &= c_1 \cdots c_k.d_1d_2d_3 \\ &\vdots \end{aligned}$$

Moreover, note that this sequence is both monotone and bounded (by $|x_0| + 1$, for instance). By the Monotone Convergence Theorem, the sequence converges to a real number, namely the number you are trying to tell me about!

For more about what it means to specify a real number, and also about why we need the real numbers rather than the rationals for calculus, read the optional Section 2.7.

Key points from Section 2.6:

- Monotone sequences (Definition 2.39)
- Monotone Convergence Theorem (Theorem 2.43)
- Principle of Mathematical Induction (page 84)
- Induction proofs with recursive sequences (Examples 2.44, 2.46, 2.48)

2.7. Optional: Completeness

The structure of Definition 2.14 suggests that, in order to prove that a sequence (z_n) converges, we need to have a candidate limit w in mind, so that we can investigate the distances $|z_n - w|$ and play the convergence game. But there is a problem: we generally do not have an explicit way of naming the limit w, except to say that it is the limit of the sequence we are investigating! In order to illustrate this point, we present two examples from the theory of integration reviewed on pages 63–66. In the first example, we *do* have an explicitly named candidate for the limit, while in the second example, we do not.

EXAMPLE 2.50. Consider the integral of the function $p(x) = x^2 + 1$ over the closed interval $[1, 2]$. If R_n is the nth Riemann sum of $p(x)$ over this interval, then

$$\int_1^2 (x^2 + 1)dx = \lim_{n\to\infty} R_n.$$

However, this is not the way we generally compute this integral. Instead, we find an antiderivative $P(x) = \frac{1}{3}x^3 + x$ and use the Fundamental Theorem of Calculus:

$$\int_1^2 (x^2+1)dx = P(2) - P(1) = \frac{8}{3} + 2 - \frac{1}{3} - 1 = \frac{10}{3}.$$

So, in this case we have an explicitly named candidate for the limit of the sequence (R_n), and we could use the definition of convergence to directly prove that the sequence does indeed converge to the rational number 10/3.

EXAMPLE 2.51. Now consider the integral of the continuous function $f(x) = \ln(\ln(x))$ over the closed interval $[3, 4]$. Again, if R_n is the nth Riemann sum of $f(x)$ over this interval, then

$$\int_3^4 \ln(\ln(x))dx = \lim_{n\to\infty} R_n.$$

You are likely tempted to use the Fundamental Theorem of Calculus to compute this integral, by finding an antiderivative. However, you may be surprised to learn that there *is* no elementary formula for an antiderivative of $f(x)$! So, in this case we cannot describe the limit any more explicitly than to say it is the number that the sequence (R_n) converges to. Here are some of the terms of the sequence (to 8 decimal places of precision).

$$\begin{aligned} R_{1000} &\approx 0.22051334 \\ R_{1001} &\approx 0.22051322 \\ R_{1002} &\approx 0.22051311 \\ R_{1003} &\approx 0.22051299 \\ R_{1004} &\approx 0.22051288 \end{aligned}$$

$$\begin{aligned} R_{1005} &\approx 0.22051276 \\ R_{1006} &\approx 0.22051265 \\ R_{1007} &\approx 0.22051253 \\ R_{1008} &\approx 0.22051242 \\ R_{1009} &\approx 0.22051230 \end{aligned}$$

This example provides a hint of why sequences are so important in mathematics: many of the things we wish to discuss (e.g., areas under curves) cannot be described in explicit, finite terms, but instead can only be named as the limits of infinite sequences. In fact, most real numbers are like this, in contrast to rational numbers. If I have a particular rational number in mind, then I have no trouble telling you exactly which one it is: I simply tell you that it is the ratio of two particular integers m/n. But if I have a particular real number in mind, how can I communicate it to you? Some numbers, like $\sqrt{2}$ and π, can be easily communicated because they have some special algebraic or geometric property: $\sqrt{2}$ is the only positive real number with square 2; the number π is the ratio circumference/diameter for any circle. But most real numbers don't have such special properties, so how do we name them?

Well, you may be thinking: just tell me the decimal expansion, and then I will know the number. True, but note that the decimal expansion of an irrational real number is infinite and non-repeating, so the best I can do is tell you more and more digits. When I do this, I am really providing a sequence of rational numbers that converges to the real number I have in mind:

$$(3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415,\ 3.14159,\ 3.141592,\ \dots).$$

is the beginning of this sequence of rational approximations of π.

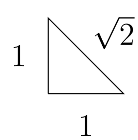
So, if real numbers are so much harder to name than rational numbers, why do we bother with them? Why not simply do calculus using rational numbers? After all, if I ask you to visualize the rational numbers, you will probably see the same mental image that you use for the reals: a line stretching to the left and right without bound. But it turns out that there are lots and lots of tiny "holes" in the rational line, despite the fact that between any two rational numbers there is another rational number (their average, for instance). These holes in the rational numbers are bad for calculus, essentially because they prevent many sequences from converging. The real numbers are exactly the result of plugging these holes, and the lack of holes in the reals makes them suitable for the development of calculus. The technical term for this property of the real numbers is *completeness*—if you later take an analysis course, you will study the concept of completeness in depth.

The most famous hole in the rational numbers is the square root of 2:

PROPOSITION 2.52. *There is no rational number m/n with the property that $(m/n)^2 = 2$.*

PROOF. We will get a contradiction by supposing that there is such a rational number m/n. We can assume that the integers m and n are both positive and share no common factors. In particular, at most one of them is even. But if $(m/n)^2 = 2$, then $m^2 = 2n^2$ and m^2 must be even. But the square of an odd number is odd, which means that m itself must be even. So we can write $m = 2k$ for some positive integer k. But then we find that $2n^2 = m^2 = (2k)^2 = 4k^2$, and dividing by 2 yields $n^2 = 2k^2$, showing that n^2 is also even. But this implies that n itself is even, contradicting our assumption that m and n share no factors. □

OK, but why does this constitute a hole in the rational numbers? After all, there is no rational number whose square is -1, but we don't think of this as a hole in the rationals. The point is that we picture positive rational numbers as distances along a line, and the Pythagorean theorem assures us that the square root of 2 is a geometrically necessary distance:



So, there is a *distance* with square 2, but that distance is not a rational number. (This discovery was extremely upsetting to the ancient Greeks.) Moreover, there *are* rational numbers with squares arbitrarily close to 2. For instance, consider the Babylonian sequence (x_n) from Examples 2.17 and 2.48, which is a rational sequence that converges to $\sqrt{2}$. By our uniqueness result (Proposition 2.24) and the fact that $\sqrt{2}$ is irrational, this means (x_n) does not converge to any rational number—there is a hole in the rational numbers where $\sqrt{2}$ should be, and the irrational real number $\sqrt{2}$ plugs that hole.

The trouble with this description of holes and completeness is that we have been forced to repeatedly mention the irrational number $\sqrt{2}$ in order to explain why the rationals have a hole where that number should be. That is, it seems that we need to know what plugs a hole in order to know that there is a hole. It would be better if we could instead detect holes directly, without reference to the eventual plug. For this, we need to introduce a new idea.

At first glance, the following definition looks very similar to the definition of convergence:

Definition 2.53. A sequence of complex numbers (z_n) is *Cauchy* if for all real numbers $d > 0$, there exists an index $N > 0$ such that $|z_n - z_m| < d$ for all indices $n, m \geq N$.

Note the key difference from Definition 2.14: there is no mention of a limit w, and instead of discussing the distances $|z_n - w|$, it speaks of the distances $|z_n - z_m|$ between different terms of the sequence. In order to unpack this new definition, we formulate it as a game.

The Cauchy Game:

(1) You go first, and you challenge me with a small distance $d > 0$.
(2) Now it is my turn. Knowing your choice of distance d, I investigate the sequence (z_n) and try to find an index N such that *all* terms with index N or greater are closer to each other than your distance d. If there is such an index N, then I announce it; if no such N exists, then I lose the game.
(3) Now it is your turn again, and you verify my choice by trying to demonstrate that for all pairs of indices n, m with $n, m \geq N$, we have $|z_n - z_m| < d$. If you can find a counterexample to this assertion, then you have shown that my N is not valid; if your demonstration succeeds, then my N is valid, and I win this round of the game.
(4) Now we return to step (1) and play another round.

To say that the sequence (z_n) is Cauchy is to say that I will win every round of the Cauchy game.

It may be helpful to compare the following imprecise but brief descriptions of our two notions:

- (Convergent) The terms of the sequence (z_n) get arbitrarily close to the limit w.
- (Cauchy) The terms of the sequence (z_n) get arbitrarily close to each other.

We have been describing holes in the rationals using the concept of convergence: for example, the Babylonian sequence of rational numbers (x_n) converges to the irrational number $w = \sqrt{2}$, and this means that there is a hole in the rationals. But (x_n) is also a Cauchy sequence, so what we have here is a Cauchy sequence of rational numbers that does not converge to a rational number. Indeed, the next proposition shows that *every* convergent sequence is Cauchy.

PROPOSITION 2.54. *Suppose that (z_n) is a convergent sequence. Then (z_n) is a Cauchy sequence.*

PROOF. Let's play the Cauchy game. You challenge me with a distance $d > 0$. Now I do something a bit clever: I imagine that you

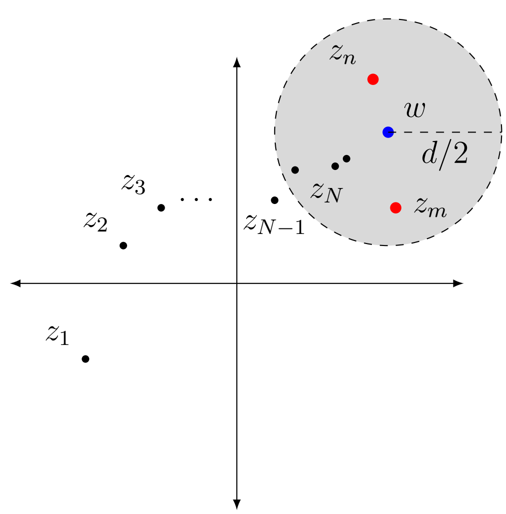


FIGURE 2.13. Convergent implies Cauchy

instead chose $d/2$. Since the sequence converges to some limit w, I can find an index N such that for all $n \geq N$, the term z_n is in the disc of radius $d/2$ centered at w (see Figure 2.13). But then for any pair of indices $n, m \geq N$, both of the terms z_n and z_m are contained in the disc of radius $d/2$ centered at w. The diameter of this disc is d, so any two points of the disc are separated by a distance less than d. In particular, we have $|z_n - z_m| < d$ as required. Since I can win in this fashion no matter what you choose for d, the sequence is Cauchy. □

The Cauchy property should appeal to your intuition in the following way: it seems plausible that if the terms of a sequence are getting arbitrarily close to each other, then there should be some particular number w that the terms are getting arbitrarily close to. That is, it seems reasonable to expect that the converse of the previous proposition should be true: if a sequence is Cauchy, then it should converge. We encourage you to think of the Cauchy property as representing a sequence's desire to converge: as you look further and further out in the sequence, the terms are huddling closer and closer together; they want to find a single point w (their limit) to huddle around. The only thing that could thwart the fulfillment of that desire is if there is a hole where w should be. So, a Cauchy sequence of rational numbers that does not converge to a rational limit signals the presence of a hole; the existence of such holes makes the rational numbers unsuitable for calculus. Note that this description of a hole does not require any mention of the real number plug that will ultimately fill the hole.

Using the notion of Cauchy sequences, we can provide a precise formulation of what we mean by the statement that the real and complex numbers have no holes:

Cauchy Completeness of $\mathbb{R}$ and $\mathbb{C}$: The real numbers $\mathbb{R}$ are *Cauchy complete*: every Cauchy sequence of real numbers converges to a real number. The complex numbers $\mathbb{C}$ are also Cauchy complete: every Cauchy sequence of complex numbers converges to a complex number.

REMARK 2.55. We are not in a position to prove that $\mathbb{R}$ and $\mathbb{C}$ are Cauchy complete; for that you will need to take an analysis course later in your mathematical education. But we have shown that the rational numbers are *not* Cauchy complete: the Babylonian sequence from Examples 2.17 and 2.48 is a Cauchy sequence of rational numbers that does not converge to a rational number.

To end this section, we will use the Cauchy completeness of $\mathbb{R}$ to prove the Monotone Convergence Theorem 2.43. Recall the statement:

Monotone Convergence Theorem: Suppose that the real sequence (x_n) is monotone and bounded. Then (x_n) converges.

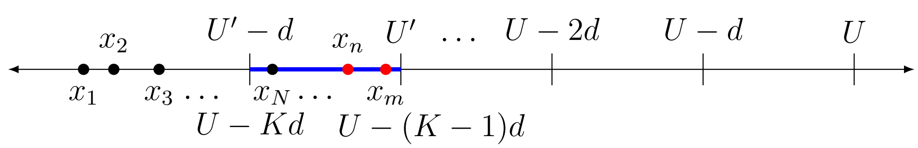
The proof closely follows the "partial proof" provided just after the statement of the MCT as Theorem 2.43 of Section 2.6.

PROOF. Assume that the sequence (x_n) is increasing (the proof of the decreasing case is similar). Let $U > 0$ be an upper bound for the sequence, so $x_n \leq U$ for all indices n. We will show that the sequence is Cauchy, hence convergent to a real number by Cauchy completeness.

So we play the Cauchy game. You begin by choosing a distance $d > 0$. I must respond with an index N. First some preliminary investigation: consider the decreasing sequence

$$(U, U-d, U-2d, U-3d, \dots).$$

I claim that there is a positive integer k such that $U-kd$ is *not* an upper bound for the sequence (x_n). This seems intuitively clear: starting at U, I am decreasing by the fixed amount $d > 0$ at each step, so eventually I must pass by one of the x's:



To justify this rigorously, we can appeal to the Archimedean property and choose $k > (U - x_1)/d$, so that

$$x_1 > U - kd.$$

Let K be the smallest such integer, so $U' = U - (K-1)d$ is an upper bound for the sequence, but $U - Kd = U' - d$ is not an upper bound. Having adjusted the upper bound in this manner, I am now ready to choose my index N.

Since $U' - d$ is not an upper bound, there exists an index N such that $x_N > U' - d$. This is my chosen index, and it is now your turn to verify my choice. So consider any two indices $n, m \geq N$. For concreteness, assume that $m \geq n \geq N$. The sequence is increasing, so we have $x_N \leq x_n \leq x_m$. But U' is an upper bound for the sequence, so putting all of this together yields

$$U' - d < x_N \leq x_n \leq x_m \leq U'.$$

This means that x_n and x_m are both contained in the half-open interval $(U' - d, U']$, which has length d. In particular, the distance between x_n and x_m is less than d. Thus, my choice of N is valid. The argument shows that I can make such a valid choice of N no matter what d you choose, and this means that the sequence is Cauchy. By Cauchy completeness of the real numbers, the sequence (x_n) converges. $\square$

Key points from Section 2.7:

- The irrationality of $\sqrt{2}$ (Proposition 2.52)
- Cauchy sequences (Definition 2.53)
- Convergent implies Cauchy (Proposition 2.54)
- Cauchy Completeness of $\mathbb{R}$ and $\mathbb{C}$ (page 97)

2.8. In-text Exercises

This section collects the in-text exercises that you should have worked on while reading the chapter.

EXERCISE 2.1 The term $z_0 = 0$ is represented by the red dot at the origin in Figure 2.1, and the next term $z_1 = c = 0.6i$ is represented by a red dot on the imaginary axis. Thinking purely geometrically (using the rotation-scale interpretation of multiplication and the parallelogram law for addition), find the red dot representing the term $z_1 = c^2 + c$. Can you identify the red dot representing z_2? What about z_3? How far can you go?

EXERCISE 2.2 This exercise asks you to consider some sequences that lie on the unit circle.

(a) Fix a positive integer $m \geq 1$, and consider the complex number $a = \cos(2\pi/m) + i\sin(2\pi/m)$. Describe the sequence consisting of the nonnegative integer powers of a:

$$(a^n) = (1,\, a,\, a^2,\, a^3,\, \dots).$$

Draw a nice picture of this sequence for $m = 6$.

(b) Now fix an irrational real number s in the interval $(0, 1)$, and define the complex number $b = \cos(2\pi s) + i\sin(2\pi s)$. Describe the sequence $(b^n) = (1\, b,\, b^2,\, b^3,\, \dots)$. What would you say is the key difference between this sequence and the sequence (a^n)?

EXERCISE 2.3 Let (w_n) denote the inverse factorial sequence from Example 2.7. Give a recursive definition for the sequence (w_n).

EXERCISE 2.4 Which of the sequences in Example 2.3 are bounded, and which are unbounded?

(a) the counting numbers (n)
(b) the harmonic sequence $(1/n)$
(c) the alternating harmonic sequence $((-1)^{n-1}/n)$
(d) the constant sequence $(i, i, i, \dots)$
(e) the sequence $(1, 2, 3, 1, 2, 3, \dots)$

(f) the sequence of prime numbers $(2, 3, 5, 7, 11, \dots)$
(g) The digits of π in base ten $(3, 1, 4, 1, 5, 9, \dots)$
(h) The iterative sequences $(0,\ c, c^2 + c, (c^2 + c)^2 + c,\ \dots)$

EXERCISE 2.5 Consider the recursively defined sequence (h_n):

$$h_1 = 1 \quad \text{and} \quad h_n = h_{n-1} + \frac{1}{n} \quad \text{for } n \geq 2.$$

(a) Compute the first 5 terms of the sequence $(h_1, h_2, h_3, h_4, h_5, \dots)$ by hand.
(b) Use a web browser to navigate to SageMathCell, located at

https://sagecell.sagemath.org

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise.

```
N = 100
tail_size = 10
h = 1.0
for n in range(2, N + 1 - tail_size):
    h = h + 1/n
for n in range(N + 1 - tail_size, N + 1):
    h = h + 1/n
    print("h_{:d} = ".format(n) + str(h))
```

Now click Evaluate. This code computes the first $N = 100$ terms of the sequence (h_n) and prints out the last tail_size $= 10$ computed terms to the screen. By changing the values of N and tail_size, you can investigate the behavior of the sequence. Based on your investigations, do you think the sequence (h_n) is bounded?

EXERCISE 2.6 Fix a positive real number $r > 0$ and an initial guess $x_{\text{guess}} > 0$ for $\sqrt{r}$. Consider the recursively defined sequence (x_n):

$$x_0 = x_{\text{guess}} \quad \text{and} \quad x_n = \frac{1}{2}\left(x_{n-1} + \frac{r}{x_{n-1}}\right) \quad \text{for } n \geq 1.$$

Use a web browser to navigate to SageMathCell, located at

https://sagecell.sagemath.org

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise in the process. Now click Evaluate. This code prints the first $N = 10$ terms of the sequence (x_n) for $r = 17$ and $x_{\text{guess}} = 4$. It then prints out the values of r and x_{N-1}^2. By changing the values of N, r, and x_{guess}, you can investigate the behavior of these sequences. Based on your investigations, do you think that (x_n) always converges to $\sqrt{r}$?

```
r = 17
x_guess = 4
N = 10
x = x_guess
print("x_0 = " + str(x))
for n in range(1, N):
    x = 0.5*(x + r/x)
    print("x_{:d} = ".format(n) + str(x))
print("\n r = " + str(r))
print("(x_{:d})^2 = ".format(n) + str(x^2))
```

Exercise 2.7 Suppose that (z_n) and (w_n) are complex sequences, and suppose that (w_n) and $(z_n + w_n)$ are both convergent sequences. Use part (b) of Proposition 2.20 to prove that (z_n) must be convergent as well.

Exercise 2.8 What is the difference between saying that the real sequence (a_n) is unbounded and saying that $\lim_{n\to\infty} a_n = \pm\infty$? Can you give an example of a positive sequence (a_n) that is unbounded but which does not diverge to ∞?

Exercise 2.9 Let $f : (0, \infty) \to \mathbb{R}$ be a real function and L a real number. In your calculus course, you may have seen the following formal definition for the statement that $\lim_{x\to\infty} f(x) = L$:

For every positive real number $d > 0$, there exists a positive real number $M > 0$ such that for all $x \geq M$, we have $|f(x) - L| < d$.

(a) Write a short paragraph explaining how this formal definition corresponds to a more intuitive understanding of the limiting behavior depicted in Figures 2.9 and 2.10:

$$\lim_{x \to \infty} \frac{1}{x} = 0 \qquad \text{and} \qquad \lim_{x \to \infty} \arctan(x) = \frac{\pi}{2}.$$

(b) Use the formal definition provided above together with Definition 2.14 to give a rigorous proof of Proposition 2.30 in the case where L is a real number.

Exercise 2.10 Use Definition 2.14 to give a careful proof of Proposition 2.36. The key point to notice is that $||z_n| - 0| = |z_n - 0|$, so playing the convergence game for $(|z_n|)$ is the same as playing the convergence game for (z_n).

Exercise 2.11 Show that a real sequence is bounded exactly when it is bounded above and bounded below.

Exercise 2.12 Suppose that U is an upper bound for a convergent real sequence (a_n), so that $a_n \leq U$ for all n. Show that $\lim_{n\to\infty} a_n \leq U$.

2.9. Problems

2.1. Write the first 6 terms of the sequence:

(a) $\left(\cos\left(\frac{\pi n}{4}\right) + i \sin\left(\frac{\pi n}{4}\right)\right)_{n\geq 1}$

(b) $\left((-i)^n \frac{1}{n+3}\right)_{n\geq 0}$

(c) $\left((-1)^n \frac{n}{n+1}\right)_{n\geq 1}$

(d) $\left(\left(\frac{-i}{2}\right)^n\right)_{n\geq 0}$

(e) $\left(\frac{n}{n!}\right)_{n\geq 1}$

2.2. What are the next three terms of the sequence? Explain.

$$(1, 11, 21, 1211, 111221, 312211, \dots).$$

2.3. Find two non-constant sequences (z_n) and (w_n) such that the sequence of products $(z_n w_n)$ is a constant sequence.

2.4. Prove that the sequence $\left(\frac{2n^2+3}{n^2+1}\right)$ is bounded.

2.5. For the following sequences and proposed limit points w, start by playing the convergence game and filling out the following table (if possible):

d = desired closeness to w	N = index to achieve d
1/2	
1/4	
1/10	
1/100	

If you believe that you can win *every* round of the convergence game, then prove that the sequence converges to w. If there exists a choice of $d > 0$ for which you will lose the convergence game, use it to prove that the sequence does not converge to the proposed limit w:

(a) $\left(\frac{1}{n+2}\right)_{n\geq 0}$ $\quad w = 0$

(b) $\left((-1)^n\right)_{n\geq 1}$ $\quad w = 1$

(c) $\left(\frac{n}{2n+1}\right)_{n\geq 1}$ $\quad w = 0$

(d) $\left(\frac{n}{2n+1}\right)_{n\geq 1}$ $w = 1$

(e) $\left(\frac{n}{2n+1}\right)_{n\geq 1}$ $w = 1/2$

(f) $\left(\cos\left(\frac{\pi n}{3}\right)\right)_{n\geq 1}$ $w = 0$

2.6. Find a sequence of real numbers (x_n) converging to 1 such that infinitely many terms are greater than 1 and infinitely many terms are less than 1.

2.7. Find a sequence of complex numbers (z_n), none of which are real, that converges to 2.

2.8. Give an example of two divergent sequences (z_n) and (w_n) such that the product sequence $(z_n w_n)$ converges.

2.9. Give an example of a divergent sequence (z_n) and a convergent sequence (w_n) such that the product sequence $(z_n w_n)$ converges. What is the limit of (w_n) in your example? Is it possible to find an example with a different limit? If so, provide such an example; if not, explain why no other limit for (w_n) is possible.

2.10. Use Propositions 2.20 and 2.30 to find the limit or explain why the sequence diverges.

(a) $\left(\cos\left(\frac{2}{n}\right)\right)$

(b) $\left(\frac{3n^3-n}{2n^3+2n+1}\right)$

(c) $\left(\frac{3-4^n}{2+3\cdot 4^n}\right)$

(d) $\left(\arctan\left(\frac{1}{n^2}\right)\right)$

(e) $\left(\frac{\ln(n+1)}{\ln(n^2)}\right)$

(f) $\left(\left(1+\frac{2}{n}\right)^n\right)$

(g) $\left(\frac{4^{n+2}}{5^n}\right)$

(h) $\left(\frac{i^n}{(0.5)^{n+1}}\right)$

(i) $(\ln(n+1) - \ln(n))$

(j) $\left(\sqrt{\frac{n^2+1}{2n^2-1}}\right)$

2.11. (The Squeeze Theorem) Suppose that $(a_n), (b_n)$, and (c_n) are real sequences satisfying $a_n \leq b_n \leq c_n$ for all n. Also suppose that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$. Prove that $\lim_{n\to\infty} b_n = L$ by completing the following steps. Let $d > 0$ be given.

(a) Show that there exists $N_1 > 0$ such that $-d < a_n - L$ for all $n \geq N_1$.

(b) Show that there exists $N_2 > 0$ such that $c_n - L < d$ for all $n \geq N_2$.

(c) Finally, show that there exists $N > 0$ such that $|b_n - L| < d$ for all $n \geq N$.

2.12. Use the Squeeze Theorem (Problem 2.11) to find the limits of the following sequences:

(a) $\left(\frac{\cos(n)}{n}\right)$

(b) $\left(\frac{4n^2+\cos(2n)}{n^2+1}\right)$

(c) $\left(2^{\frac{1}{n}}\right)$ [Hint: use the binomial expansion from Example 2.8 to show that $2 \leq (1 + \frac{1}{n})^n$ for all n]

(d) $\left((1 + \frac{n}{n+1})^{\frac{1}{n}}\right)$ [Hint: use part (c)]

2.13. Determine whether the following sequences are bounded, monotone, both or neither.

(a) $\left(n + \frac{1}{n}\right)$

(b) $\left(ne^{-n}\right)$

(c) $\left(\frac{n}{n+1}\right)$

(d) $\left(\frac{(-1)^n}{2n+3}\right)$

2.14. Consider the recursive sequence defined by:

$$a_1 = \sqrt{2} \quad \text{and} \quad a_n = \sqrt{2a_{n-1}} \quad \text{for } n \geq 2.$$

(a) Write out the first 4 terms of the sequence.

(b) Show that $a_1 \leq a_2$.

(c) Use induction to show that the sequence (a_n) is increasing.

(d) Show that $a_1 \leq 2$.

(e) Use induction to show that the sequence (a_n) is bounded by 2.

(f) Find the limit of (a_n).

2.15. Consider the recursive sequence defined by:

$$a_1 = \sqrt{2} \quad \text{and} \quad a_n = \sqrt{2 + a_{n-1}} \quad \text{for } n \geq 2.$$

(a) Write out the first 4 terms of the sequence.

(b) Show that $a_1 \leq a_2$.

(c) Use induction to show that the sequence (a_n) is increasing.

(d) Show that $a_1 \leq 2$.
(e) Use induction to show that the sequence (a_n) is bounded by 2.
(f) Find the limit of (a_n).

2.16. Consider the recursive sequence defined by:

$$a_1 = 1 \quad \text{and} \quad a_n = 3 - \frac{1}{a_{n-1}} \quad \text{for } n \geq 2.$$

(a) Write out the first 4 terms of the sequence.
(b) Show that $a_1 \leq a_2$.
(c) Use induction to show that the sequence (a_n) is increasing.
(d) Show that $a_1 \leq 3$.
(e) Use induction to show that the sequence (a_n) is bounded by 3.
(f) Find the limit of (a_n).

2.17. Let (z_n) be a sequence of complex numbers that converges to 0. Let (c_n) be a bounded sequence of complex numbers. Show that $(c_n z_n)$ converges to 0.

2.18. This problem leads you to a combinatorial interpretation of the binomial coefficients $\binom{n}{k}$. Consider the set containing the first n positive integers:

$$\{1, 2, 3, \cdots, n\}.$$

For each integer k with $0 \leq k \leq n$, let $b(n, k)$ denote the number of subsets of this set containing exactly k numbers; the order of the numbers does not matter, only which ones are present in the subset. For example, taking $n = 3$, we are looking at the subsets of $\{1, 2, 3\}$:

$k = 0:$	$\{\,\} =$ empty set	$b(3, 0) = 1$
$k = 1:$	$\{1\}, \{2\}, \{3\}$	$b(3, 1) = 3$
$k = 2:$	$\{1, 2\}, \{1, 3\}, \{2, 3\}$	$b(3, 2) = 3$
$k = 3:$	$\{1, 2, 3\}$	$b(3, 3) = 1$

Note that $b(n, k)$ is the number of ways of choosing k elements from a set with n elements.

(a) Carefully explain why the number $b(3,k)$ appears as the coefficient of z^k in the expansion of the polynomial

$$(1+z)^3 = (1+z)(1+z)(1+z).$$

Hint: identify the first, second, and third linear factor written above with the numbers $1,2,3$ and think about the definition of $b(3,k)$ as you expand the product.

(b) Now explain why, for every $n \geq 1$, the number $b(n,k)$ appears as the coefficient of z^k in the expansion of the polynomial

$$(1+z)^n = (1+z)(1+z)\cdots(1+z).$$

You have now shown that $b(n,k) = \binom{n}{k}$, which explains the name "n choose k" for the binomial coefficients.

2.19. Let D be the open disc of radius $d > 0$ centered at a complex number w:

$$D = \{z \text{ in } \mathbb{C} : |z-w| < d\}.$$

Fix a point a in the disc D, and consider the open disc D' of radius $d' = d - |a-w| > 0$ centered at a:

$$D' = \{z \text{ in } \mathbb{C} : |z-a| < d'\}.$$

Draw a nice picture illustrating this situation. Then use the triangle inequality (Proposition 1.7) to show that D' is entirely contained inside the original disc D.

2.20. Consider the sequence (x_n) with nth term $x_n = \left(1+\frac{1}{n}\right)^n$. Use a web browser to navigate to SageMathCell, located at

`https://sagecell.sagemath.org`

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise in the process. Now click Evaluate. This code prints the first $N = 10$ terms of the sequence (x_n). By changing the value of N, you can investigate the behavior of the sequence. Based on your investigations, do you think that (x_n) converges? If so, do you have a guess for the limit?

```
N = 10
for n in range(1, N+1):
    x = (1+1/n)**n
    print("x_{:d} = ".format(n) + str(x.n()))
```

2.21. Adapt the proof from Example 2.48 to show that the sequence from Exercise 2.6 converges to $\sqrt{r}$:

$$x_0 = 1 \quad \text{and} \quad x_n = \frac{1}{2}\left(x_{n-1} + \frac{r}{x_{n-1}}\right) \quad \text{for } n \geq 1.$$

2.22. Fix a positive real number $r > 0$. Consider the recursively defined sequence (y_n):

$$y_0 = 1 \quad \text{and} \quad y_n = \frac{1}{3}\left(2y_{n-1} + \frac{r}{y_{n-1}^2}\right) \quad \text{for } n \geq 1.$$

Use a web browser to navigate to SageMathCell, located at

https://sagecell.sagemath.org

Copy and paste the Python code provided below into the window, being careful to fix any indentation problems that may arise in the process. Now click Evaluate. This code prints the first $N = 10$ terms of the sequence (y_n) for $r = 2$. By changing the values of r and N, you can investigate the behavior of these sequences. Based on your investigations, do you think that (y_n) always converges? If so, do you have a guess for the limit?

```
r = 2
N = 10
y = 1.0
print("y_0 = " + str(y))
for n in range(1, N):
    y = (1/3)*(2*y + r/(y**2))
    print("y_{:d} = ".format(n) + str(y))
```

CHAPTER 3

SERIES

3.1. Inspiration: The Riemann Zeta Function

By way of motivation for this chapter, we now introduce one of the most interesting complex functions in all of mathematics: the *Riemann zeta function*, $\zeta(z)$. Figure 3.1 shows the graph of the magnitude $|\zeta(z)|$ on the domain $\mathrm{Re}(z) > 1$. This fascinating function arises in the subject of number theory and has applications to other areas of mathematics and physics. The Riemann Hypothesis, formulated by the German mathematician Bernhard Riemann in 1859, concerns the location of the zeros of the zeta function and has profound implications for the distribution of prime numbers among all the integers. In 2000, the Clay Mathematics Institute listed the Riemann Hypothesis as one of its seven Millennium Prize Problems, each of which carries a one million dollar prize. The Riemann Hypothesis is still open, so pay attention: the ideas in this chapter could set you on the path to a lucrative reward.

Our reason for introducing the zeta function is because its definition requires the notion of *infinite series*, the central topic of this chapter.

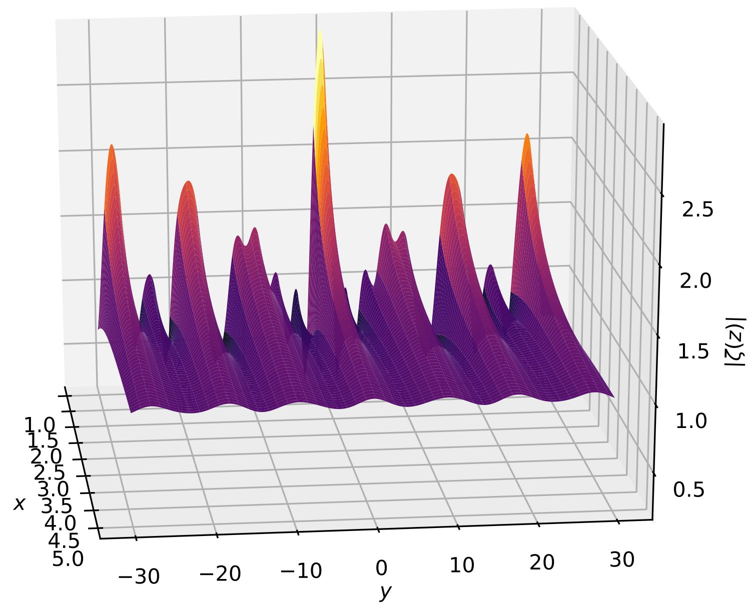


FIGURE 3.1. Graph of the magnitude of the Riemann zeta function $|\zeta(z)|$. Note that the axes are rotated from the usual representation, with the imaginary y-axis running horizontally across the page.

Here is the defining formula:

$$\zeta(z) = 1 + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \cdots .$$

Right away, there are at least two things about this formula that need discussion:

(1) The variable z is complex, so we seem to be contemplating complex number exponents. But what does 2^z mean? We will return to this question in the optional Section 4.10, as an application of the complex exponential function studied in Section 4.7. (Recall that we encountered a complex extension of the real exponential function e^x in Example 1.20.) For now, we will restrict attention to real number exponents p, for which

the zeta function has the formula

$$\zeta(p) = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \frac{1}{4^p} + \cdots .$$

Here, we are thinking of p as a real variable, so we will want to investigate these expressions for various specific values of p, for instance $p = 2$ and $p = \pi$:

$$\begin{aligned} \zeta(2) &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \\ \zeta(\pi) &= 1 + \frac{1}{2^\pi} + \frac{1}{3^\pi} + \frac{1}{4^\pi} + \cdots . \end{aligned}$$

(2) The dots "$\cdots$" indicate that the sum goes on forever. So we need to make sense of the activity of adding up an infinite sequence of numbers. In this case, the infinite sequence is $(1/n^p) = (1, 1/2^p, 1/3^p, 1/4^p, \dots)$, but we may investigate this notion of "infinite summation" for any sequence of complex numbers (c_n).

3.2. Series and Convergence

Suppose that (c_n) is a sequence of complex numbers. We want to make sense of "adding up all of the numbers c_n." Your initial reaction to this idea may be that the process will never end, or if it does somehow come to an end, then the result must be infinite or nonsensical. After all, if there are really infinitely many terms c_n, then there are always more terms to add. This conundrum should remind you of the improper integrals you studied in MATH 155, where you discovered that some regions having infinite extent actually have only finite area (we will return to improper integrals in Section 3.3).

Luckily, our notion of convergence saves the day. Instead of prejudging the outcome of "infinite summation," we proceed slowly and investigate what happens as we add one additional term at a time to obtain the sequence of finite partial sums (s_m):

$$\begin{aligned} s_1 &= c_1 \\ s_2 &= c_1 + c_2 \\ s_3 &= c_1 + c_2 + c_3 \\ &\vdots \\ s_m &= c_1 + c_2 + c_3 + \cdots + c_m \\ &\vdots \end{aligned}$$

Having done this, it makes sense to ask whether the sequence (s_m) converges or not. If it *does* converge to a finite limit s, then we should think of s as the sum of *all* of the numbers c_n, because the partial sums s_m get arbitrarily close to s as we include more and more of the terms c_n. On the other hand, if the sequence (s_m) *does not* converge, then we should conclude that it is not possible to add up all of the terms c_n, there being no single finite value that the partial sums s_m are approaching.

We record these ideas in the following definition.

DEFINITION 3.1. Let $(c_n)_{n\geq 1}$ be a sequence of complex numbers. Define a new sequence of complex numbers $(s_m)_{m\geq 1}$, the sequence of *partial sums*, as follows:

$$s_m = c_1 + c_2 + c_3 + \cdots + c_m = \sum_{n=1}^{m} c_n.$$

In words, for each index $m \geq 1$, the mth partial sum s_m is the finite sum of the first m terms of the sequence (c_n). If the sequence of partial sums (s_m) converges to a limit s, then we say that the *series* $c_1+c_2+c_3+\cdots$ *converges to the sum* s, and we write

$$\sum_{n=1}^{\infty} c_n = c_1 + c_2 + c_3 + \cdots = \lim_{m\to\infty} s_m = s.$$

If the sequence of partial sums (s_m) does not converge, then we say that the series $\sum_{n=1}^{\infty} c_n$ *diverges*.

REMARK 3.2. If the original sequence $(c_n)_{n\geq 0}$ starts with the index 0 or some other integer, then the indices of the sequence of partial sums $(s_m)_{m\geq 0}$ are adjusted accordingly:

$$\begin{aligned} s_0 &= c_0 \\ s_1 &= c_0 + c_1 \\ &\vdots \\ s_m &= c_0 + c_1 + \cdots + c_m = \sum_{n=0}^{m} c_n. \end{aligned}$$

In this case, we write $\sum_{n=0}^{\infty} c_n = \lim_{m\to\infty} s_m$ when the series converges.

The best way to come to grips with this definition is to start working with examples. At the end of this section (Remark 3.14), we provide some general comments about common sources of mistakes when working with series; those comments are probably best appreciated after having made some of the mistakes for yourself.

EXAMPLE 3.3. Consider the sequence $(c_n) = \left(\frac{1}{n(n+1)}\right)$:

$$c_1 = \frac{1}{1\cdot 2} = \frac{1}{2}, \quad c_2 = \frac{1}{2\cdot 3} = \frac{1}{6}, \quad c_3 = \frac{1}{3\cdot 4} = \frac{1}{12}, \quad \ldots.$$

We wish to investigate the series

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \cdots.$$

Does it converge or diverge, and if it does converge, can we name the sum explicitly? Let's begin by looking numerically (to 6 decimal places of precision) at the sequence of partial sums:

$$\begin{aligned} s_1 &= 0.500000 & &\vdots \\ s_2 &\approx 0.666667 & s_{1000} &\approx 0.999001 \\ s_3 &= 0.750000 & s_{1001} &\approx 0.999002 \\ s_4 &= 0.800000 & s_{1002} &\approx 0.999003 \\ s_5 &\approx 0.833333 & s_{1003} &\approx 0.999004 \\ & & s_{1004} &\approx 0.999005 \end{aligned}$$

Could it be that these partial sums are converging to 1? Yes, and if we use a bit of algebra, we can prove it. Note that we can rewrite the numbers c_n as follows:

$$c_n = \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

This allows us to compute the partial sums exactly:

$$\begin{aligned} s_m &= c_1 + c_2 + c_3 + \cdots + c_m \\ &= \left(1 - \cancel{\frac{1}{2}}\right) + \left(\cancel{\frac{1}{2}} - \cancel{\frac{1}{3}}\right) + \left(\cancel{\frac{1}{3}} - \cancel{\frac{1}{4}}\right) + \cdots + \left(\cancel{\frac{1}{m}} - \frac{1}{m+1}\right) \\ &= 1 - \frac{1}{m+1}. \end{aligned}$$

Note the fortuitous cancelation in the second line, in which the negative term in each parenthesis is canceled by the positive term in the next parenthesis. Because of this phenomenon, series like this one are called *telescoping*, a name derived from the fact that some telescopes are made of segments that slide inside of each other to make the instrument short when not in use.

Because of the telescoping behavior, we have found a simple formula for the mth partial sum of the series:

$$s_m = 1 - \frac{1}{m+1}.$$

Moreover, since the harmonic sequence $(1/(m+1))$ converges to zero, an application of Proposition 2.20(b) shows that

$$\lim_{m\to\infty} s_m = \lim_{m\to\infty}\left(1 - \frac{1}{m+1}\right) = 1 - \lim_{m\to\infty}\frac{1}{m+1} = 1.$$

Hence, we are justified in saying that the series converges to 1, and writing

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

REMARK 3.4. Be careful with the words *sequence* and *series*: the symbol $\sum_{n=1}^{\infty} c_n$ denotes a *series*. This series has two important *sequences* associated with it: the sequence of terms (c_n) that we are attempting to sum, and the sequence of partial sums (s_m). Note in particular that $\lim_{n\to\infty} c_n$ denotes the limit of the sequence of terms (c_n), which is quite different from the infinite series $\sum_{n=1}^{\infty} c_n = \lim_{m\to\infty} s_m$.

For instance, in the previous example of the series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$, the sequence of terms is $(c_n) = \left(\frac{1}{n(n+1)}\right)$ and the sequence of partial sums is $(s_m) = (1 - \frac{1}{m+1})$. So we have

$$\lim_{n\to\infty} c_n = \lim_{n\to\infty} \frac{1}{n(n+1)} = 0,$$

while

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{m\to\infty} s_m = \lim_{m\to\infty} \left(1 - \frac{1}{m+1}\right) = 1.$$

EXAMPLE 3.5. Consider the constant sequence $(c_n) = (1+i)_{n\geq 1}$, and let's investigate the series $\sum_{n=1}^{\infty}(1+i)$. As always, we begin by looking at the partial sums:

$$s_m = (1+i) + (1+i) + \cdots + (1+i) = m + mi.$$

The sequence (s_m) is not bounded, because

$$|s_m| = |m + mi| = \sqrt{m^2 + m^2} = m\sqrt{2}.$$

By Proposition 2.18, the sequence of partial sums (s_m) does not converge, so the series $\sum_{n=1}^{\infty}(1+i)$ diverges. In this case, the divergence of the series is due to the sequence of partial sums (s_m) being unbounded.

EXAMPLE 3.6. Consider the alternating sequence

$$(c_n) = ((-1)^{n+1})_{n\geq 1} = (1, -1, 1, -1, 1, -1, \dots).$$

We wish to investigate the series

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots = \sum_{n=1}^{\infty} (-1)^{n+1}.$$

We look at the partial sums:

$$\begin{aligned} s_1 &= 1 \\ s_2 &= 1 - 1 = 0 \\ s_3 &= 1 - 1 + 1 = 1 \\ s_4 &= 1 - 1 + 1 - 1 = 0 \\ &\vdots \end{aligned}$$

We see that the partial sums form the sequence $(s_m) = (1, 0, 1, 0, 1, 0, \dots)$. This sequence does not converge, so the series $\sum_{n=1}^{\infty}(-1)^{n+1}$ diverges. Note that this type of divergence is a bit different from the previous example, where the partial sums formed an unbounded sequence. In this case, even though the sequence of partial sums (s_m) is bounded, it does not settle down (converge) to a single value, and hence the series diverges.

Exercise 3.1. What is wrong with the following telescoping argument that purports to show that the series sums to 0?

$$\begin{aligned} \sum_{n=1}^{\infty}(-1)^{n+1} &= 1 - 1 + 1 - 1 + 1 - 1 + \cdots \\ &= (1 - 1) + (1 - 1) + (1 - 1) + \cdots \\ &= 0 + 0 + 0 + \cdots \\ &= 0. \end{aligned}$$

Can you make a similar (invalid) argument that suggests that the series sums to 1?

The previous two examples of divergent series are perhaps not too surprising in hindsight: in both cases the terms c_n do not get arbitrarily small as the index n gets bigger and bigger. But the terms c_n are exactly the changes in the partial sums:

$$s_m = c_1 + c_2 + \cdots + c_{m-1} + c_m = s_{m-1} + c_m.$$

So the fact that the numbers c_m do not approach zero means that the terms s_m continue to change substantially as m gets bigger and bigger, and this means that the sequence (s_m) cannot converge. The next result makes these observations precise.

PROPOSITION 3.7 (Divergence Test). *Suppose that the series* $\sum_{n=1}^{\infty} c_n$ *converges to the sum* s. *Then the sequence of terms* (c_n) *converges to zero.*

PROOF. We are assuming that $\sum_{n=1}^{\infty} c_n = s$. But this just means that the sequence of partial sums (s_m) converges to s:

$$\lim_{m\to\infty} s_m = s.$$

Observe that we can also write $\lim_{m\to\infty} s_{m-1} = s$. Indeed, the sequence (s_{m-1}) is the *same* list of numbers in the *same* order, only with the indices shifted by 1.

But, as explained just above, we have $(c_m) = (s_m - s_{m-1})$, so an application of Proposition 2.20(b) shows that

$$\lim_{m\to\infty} c_m = \lim_{m\to\infty} (s_m - s_{m-1}) = s - s = 0,$$

so the sequence of terms (c_m) converges to zero as claimed. □

REMARK 3.8. The previous proposition goes by the name of *the divergence test*, because it is most often used to show that a series is divergent. We explain this in two steps:

(1) The statement of the proposition (which mentions only convergence) is logically equivalent to a statement that mentions only divergence:

$$\sum_{n=1}^{\infty} c_n \quad \text{converges} \qquad \textbf{implies} \qquad \lim_{n\to\infty} c_n = 0$$

is logically equivalent to

$$\lim_{n\to\infty} c_n \neq 0 \qquad \textbf{implies} \qquad \sum_{n=1}^{\infty} c_n \quad \text{diverges.}$$

(2) We put this reformulation to work in the following way: given a series $\sum_{n=1}^{\infty} c_n$, it is a good idea to first investigate the sequence of terms (c_n). If we find that (c_n) does not converge or that the limit exists but is different from zero, we can immediately conclude that the series diverges. If, on the other hand, we find that $\lim_{n\to\infty} c_n = 0$, then we must move on to an investigation of the sequence of partial sums (s_m) to determine if the series converges or diverges.

Remember: the divergence test can only demonstrate divergence, never convergence. It may be helpful to think of it this way: there is no hope for a series $\sum_{n=1}^{\infty} c_n$ to converge unless the sequence of terms (c_n) converges to zero; but just because the sequence (c_n) converges to zero does not in itself guarantee that the series converges. We will see some examples of this in the next section.

Example 3.9. (Divergent p-series) Consider the sequence

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

This sequence is unbounded, and so in particular $\lim_{n\to\infty} c_n \neq 0$. The series $\sum_{n=1}^{\infty} \sqrt{n}$ diverges by the divergence test.

More generally, consider the p-sequence $(1/n^p)$. For $p < 0$, the sequence is unbounded (Example 2.34), while for $p = 0$ the sequence $(1/n^0) = (1)_{n\geq 1}$ is a nonzero constant. By the divergence test, the *p-series* $\sum_{n=1}^{\infty} 1/n^p$ diverges whenever the exponent $p \leq 0$.

Note that for $p > 0$, the p-sequence $(1/n^p)$ converges to zero, and the divergence test cannot help us: the corresponding p-series $\sum_{n=1}^{\infty} 1/n^p$ may or may not converge, and we will use other methods to determine the truth in Proposition 3.18 of the next section.

Example 3.10 (Geometric Series). Consider the geometric sequence with first term $a = 2$ and common ratio $c = 3i$:

$$(ac^n)_{n\geq 0} = (2 \cdot (3i)^n)_{n\geq 0} = (2,\, 6i,\, -18,\, -54i,\, 162,\, \ldots).$$

This sequence is unbounded, because the magnitude of the common ratio $|c| = |3i| = 3 > 1$. Hence the geometric series $\sum_{n=0}^{\infty} 2 \cdot (3i)^n$ diverges by the divergence test.

More generally, fix two complex numbers $a \neq 0$ and c, and consider the geometric sequence $(ac^n)_{n\geq 0}$. As noted above, this sequence is unbounded if $|c| > 1$ by Example 2.35:

$$\lim_{n\to\infty} |ac^n| = \lim_{n\to\infty} |a||c|^n = |a| \lim_{n\to\infty} |c|^n = +\infty.$$

If $|c| = 1$, then the sequence (ac^n) lies on the circle of radius $|a| \neq 0$, and hence does not converge to zero. By the divergence test, the geometric series $\sum_{n=0}^{\infty} ac^n$ diverges whenever $|c| \geq 1$.

Now assume that $|c| < 1$, so the sequence $(|c|^n)$ converges to zero, which implies that the sequence (ac^n) also converges to zero:

$$\lim_{n\to\infty} ac^n = a \lim_{n\to\infty} c^n = a \cdot 0 = 0.$$

Remember: the divergence test now only tells us that it is *possible* for the series $\sum_{n=0}^{\infty} ac^n$ to converge, but cannot help us further to actually establish the convergence. For that, we will need to take a different approach, beginning with the following telescoping computation with polynomials:

$$\begin{aligned}(1-z)(1+z+z^2+\cdots+z^m) &= 1 + \cancel{z} + \cancel{z^2} + \cdots + \cancel{z^m} \\ &\qquad - \cancel{z} - \cancel{z^2} - \cdots - \cancel{z^m} - z^{m+1} \\ &= 1 - z^{m+1}.\end{aligned}$$

Dividing by $1 - z$ reveals an important algebraic identity:

$$1 + z + z^2 + \cdots + z^m = \frac{1 - z^{m+1}}{1 - z}.$$

This is an identity of complex functions, which means that the equality continues to hold after plugging in any complex number $c \neq 1$ (so as to avoid a zero in the denominator on the right hand side):

$$1 + c + c^2 + \cdots + c^m = \frac{1 - c^{m+1}}{1 - c}.$$

In particular, we can use this result to express each of the partial sums s_m of our geometric series in a compact form:

$$s_m = a(1 + c + c^2 + \cdots + c^m) = \frac{a(1 - c^{m+1})}{1 - c}.$$

Recall that, since $|c| < 1$, we have $\lim_{n\to\infty} c^n = 0$. Now we take the limit of the partial sums to find that the series converges:

$$\sum_{n=0}^{\infty} ac^n = \lim_{m\to\infty} s_m = \lim_{m\to\infty} \frac{a(1 - c^{m+1})}{1 - c} = \frac{a}{1 - c}.$$

As a concrete example of a convergent geometric series, take $a = 1$ and $c = \frac{i}{2}$. Then we have

$$\sum_{n=0}^{\infty} \left(\frac{i}{2}\right)^n = \frac{1}{1 - \frac{i}{2}} = \frac{2}{2 - i} = \frac{2}{5}(2 + i).$$

Figure 3.2 shows both the sequence of terms $((i/2)^n)$ and the sequence of partial sums $(s_m) = \left(\sum_{n=0}^{m}(i/2)^n\right)$.

We now prove three important limit laws for series. In fact, these results are immediate consequences of the corresponding facts for sequences. (Proposition 2.20).

PROPOSITION 3.11 (Limit Laws for Series). *Suppose that the series $\sum_{n=1}^{\infty} c_n$ converges to the sum s and the series $\sum_{n=1}^{\infty} d_n$ converges to the sum t. That is, we have*

$$\sum_{n=1}^{\infty} c_n = s \qquad \textit{and} \qquad \sum_{n=1}^{\infty} d_n = t.$$

Then

(a)

$$\sum_{n=1}^{\infty}(c_n + d_n) = \sum_{n=1}^{\infty} c_n + \sum_{n=1}^{\infty} d_n = s + t;$$

(b)

$$\sum_{n=1}^{\infty}(c_n - d_n) = \sum_{n=1}^{\infty} c_n - \sum_{n=1}^{\infty} d_n = s - t;$$

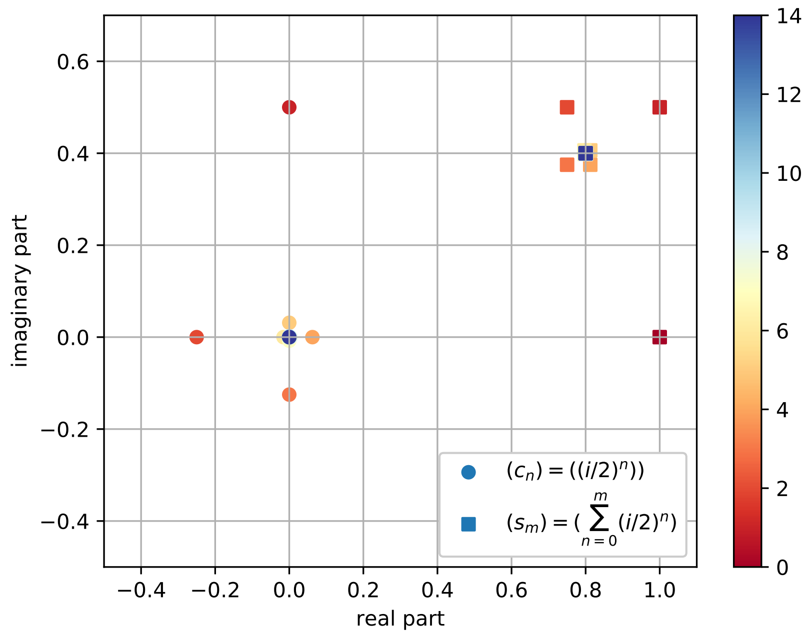


FIGURE 3.2. The sequence of terms and the sequence of partial sums for the geometric series $\sum_{n=0}^{\infty}(i/2)^n$. The first 15 terms of the sequence $((i/2)^n)$ are displayed as round dots; the first 15 terms of the sequence of partial sums (s_m) are displayed as squares.

(c) if a is any complex number, then

$$\sum_{n=1}^{\infty} ac_n = a\sum_{n=1}^{\infty} c_n = as.$$

PROOF. We will prove part (a) and leave the other parts as exercises. Let (s_m) denote the sequence of partial sums of the series $\sum_{n=1}^{\infty} c_n$, and (t_m) the sequence of partial sums of $\sum_{n=1}^{\infty} d_n$. Then we are assuming that (s_m) converges to s and (t_m) converges to t. But consider the sequence of partial sums (w_m) of the series $\sum_{n=1}^{\infty}(c_n + d_n)$:

$$\begin{aligned} w_m &= (c_1 + d_1) + (c_2 + d_2) + \cdots + (c_m + d_m) \\ &= (c_1 + c_2 + \cdots + c_m) + (d_1 + d_2 + \cdots + d_m) \\ &= s_m + t_m. \end{aligned}$$

It follows by Proposition 2.20(a) that

$$\lim_{m\to\infty} w_m = \lim_{m\to\infty}(s_m + t_m) = s + t.$$

This means that the series $\sum_{n=1}^{\infty}(c_n + d_n)$ converges to $s+t$ as claimed. □

EXERCISE 3.2. Prove parts (*b*) and (*c*) of Proposition 3.11.

We now state a series analogue of Proposition 2.25 that relates the convergence of a complex sequence to the convergence of its real and imaginary parts.

PROPOSITION 3.12. *Suppose that $\sum_{n=1}^{\infty} c_n$ is a complex series, with $(c_n) = (a_n + b_n i)$. Consider the corresponding series of real and imaginary parts $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$. Then*

$$\sum_{n=1}^{\infty} c_n = s \quad \textit{if and only if} \quad \sum_{n=1}^{\infty} a_n = \mathrm{Re}(s) \quad \textit{and} \quad \sum_{n=1}^{\infty} b_n = \mathrm{Im}(s).$$

PROOF. EXERCISE 3.3 □

EXERCISE 3.3. Use Proposition 2.25 to prove Proposition 3.12.

EXAMPLE 3.13. Here is an example of using Proposition 3.12. Consider the series

$$\sum_{n=0}^{\infty}\left(\frac{1}{2^n} + \frac{i}{3^n}\right).$$

Here, the nth term of the sequence is $c_n = \frac{1}{2^n} + \frac{i}{3^n} = a_n + b_n i$ with real part $a_n = \frac{1}{2^n}$ and imaginary part $b_n = \frac{1}{3^n}$. The corresponding real series are convergent geometric:

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty}\frac{1}{2^n} = \frac{1}{1-\frac{1}{2}} = 2$$

and

$$\sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty}\frac{1}{3^n} = \frac{1}{1-\frac{1}{3}} = \frac{3}{2}.$$

By Proposition 3.12, it follows that $\sum_{n=0}^{\infty} c_n$ is convergent, with

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty} \left(\frac{1}{2^n} + \frac{i}{3^n}\right) = \sum_{n=0}^{\infty} \frac{1}{2^n} + i \sum_{n=0}^{\infty} \frac{1}{3^n} = 2 + \frac{3}{2} i.$$

Note that the series $\sum_{n=0}^{\infty} c_n$ is not itself a geometric series, although its real part $\sum_{n=0}^{\infty} a_n$ and its imaginary part $\sum_{n=0}^{\infty} b_n$ are both geometric series.

Remark 3.14. Here are some common sources of mistakes for students beginning to study infinite series:

(a) (Notation) The finite summation symbol $\sum_{n=1}^{m} c_n$ is simply shorthand for the finite sum $c_1 + c_2 + \cdots + c_m$. This straightforward algebraic procedure always leads to a finite answer, and the usual laws of algebra always apply. The infinite summation symbol $\sum_{n=1}^{\infty} c_n$ is chosen to look a lot like the finite symbol, but its interpretation is more subtle. We encourage you to think of the symbol $\sum_{n=1}^{\infty} c_n$ not as a straightforward algebraic operation, but rather as shorthand for the investigative procedure involving the sequence of partial sums (s_m) described in Definition 3.1. Depending on the outcome of that procedure, it may be that the symbol comes to denote a specific complex number $s = \lim_{m\to\infty} s_m$ (this happens when the series converges), or it may be that the symbol comes to represent a type of nonexistence (when the series diverges, so there is no sum).

(b) (Language) We repeat here the content of Remark 3.4: be very careful in your use of the words *sequence* and *series*: the symbol $\sum_{n=1}^{\infty} c_n$ denotes a *series*. This series has two important *sequences* associated with it: the sequence of terms (c_n) that we are attempting to sum, and the sequence of partial sums (s_m). Note in particular that $\lim_{n\to\infty} c_n$ denotes the limit of the sequence of terms (c_n), which is quite different from the infinite series $\sum_{n=1}^{\infty} c_n = \lim_{n\to\infty} s_n$. See Figure 3.2 for a visual illustration of the distinction, and be sure not to confuse these two different limits.

(c) (Algebra) You should use caution when making algebraic manipulations with series, because not all of the familiar rules apply. For instance, in the optional Section 3.8, we explain how adding up the terms of a series in a different order can change the sum! To be safe, when working with infinite series you should carefully apply the limit laws (Proposition 3.11) or else manipulate the finite partial sums and then draw conclusions about the infinite sum. Refer back to Example 3.6 and Exercise 3.1 for instances of these types of mistakes.

Key points for Section 3.2:

- Definition of sequence of partial sums, series, convergence and divergence (Definition 3.1)
- Divergence Test (Proposition 3.7)
- Geometric series, convergence and divergence (Example 3.10)
- Limit laws for series (Proposition 3.11)

3.3. p-Series and the Integral Test ($\mathbb{R}$)

This section takes place entirely in the context of the real numbers $\mathbb{R}$.

In Example 3.9 of the previous section, we used the divergence test to show that the *p-series* $\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges for $p \leq 0$. We now wish to investigate the convergence or divergence of p-series for $p > 0$. The divergence test provides no help, because $\lim_{n\to\infty} 1/n^p = 0$ for $p > 0$. In order to develop some intuition, we begin by looking numerically at some specific examples. The actual proofs of convergence or divergence will come later in this section, after we develop the appropriate theoretical tool for the job.

Example 3.15. Consider the exponent $p = 2$ and corresponding p-series $\sum_{n=1}^{\infty} 1/n^2$. Denoting the sequence of partial sums by (s_m) as usual, we have (to 6 decimal places of precision)

$$\begin{aligned} s_{991} &\approx 1.643925 & s_{996} &\approx 1.643931 \\ s_{992} &\approx 1.643927 & s_{997} &\approx 1.643932 \\ s_{993} &\approx 1.643928 & s_{998} &\approx 1.643933 \\ s_{994} &\approx 1.643929 & s_{999} &\approx 1.643934 \\ s_{995} &\approx 1.643930 & s_{1000} &\approx 1.643935 \end{aligned}$$

These computations suggest that the series $\sum_{n=1}^{\infty} 1/n^2$ converges to a sum $s \approx 1.64$.

EXAMPLE 3.16. Now consider the exponent $p = \frac{1}{2}$ and the p-series $\sum_{n=1}^{\infty} 1/n^{1/2} = \sum_{n=1}^{\infty} 1/\sqrt{n}$. Here are some terms of the sequence of partial sums (s_m):

$$\begin{aligned} s_{991} &\approx 61.52 & s_{996} &\approx 61.67 \\ s_{992} &\approx 61.55 & s_{997} &\approx 61.71 \\ s_{993} &\approx 61.58 & s_{998} &\approx 61.74 \\ s_{994} &\approx 61.61 & s_{999} &\approx 61.77 \\ s_{995} &\approx 61.64 & s_{1000} &\approx 61.80 \end{aligned}$$

It is hard to know what to think at this point, so we compute some later terms of the sequence of partial sums:

$$\begin{aligned} s_{9991} &\approx 198.45 & s_{9996} &\approx 198.50 \\ s_{9992} &\approx 198.46 & s_{9997} &\approx 198.51 \\ s_{9993} &\approx 198.47 & s_{9998} &\approx 198.52 \\ s_{9994} &\approx 198.48 & s_{9999} &\approx 198.53 \\ s_{9995} &\approx 198.49 & s_{10000} &\approx 198.54 \end{aligned}$$

It appears that the sequence (s_m) may be unbounded, and thus that the series $\sum_{n=1}^{\infty} 1/\sqrt{n}$ may be divergent. As one further piece of evidence in that direction, the millionth partial sum of this series has the approximate value $s_{1000000} \approx 1998.54$.

EXAMPLE 3.17. As a final numerical example, consider the exponent $p = 1$, and the corresponding *harmonic series* $\sum_{n=1}^{\infty} 1/n$. Denoting the sequence of partial sums of this series by (h_m), we have

$$\begin{aligned} h_{990} &\approx 7.475 & h_{995} &\approx 7.480 \\ h_{991} &\approx 7.476 & h_{996} &\approx 7.481 \\ h_{992} &\approx 7.477 & h_{997} &\approx 7.482 \\ h_{993} &\approx 7.478 & h_{998} &\approx 7.483 \\ h_{994} &\approx 7.479 & h_{999} &\approx 7.484 \end{aligned}$$

In fact, you have already considered the sequence (h_m) in EXERCISE 2.5 of Section 2.3, where we introduced it recursively as

$$h_1 = 1 \quad \text{and} \quad h_m = h_{m-1} + \frac{1}{m} \quad \text{for } m \geq 2.$$

In that exercise, we asked you to investigate the sequence numerically and make a guess as to whether it is bounded or unbounded. Do you remember your answer? It is likely that you guessed the sequence to be bounded. After all, the millionth partial sum is only $h_{1000000} \approx 14.393$.

Actually, the sequence (h_m) is unbounded, so the harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges. This is an astounding fact: even though the millionth partial sum is smaller than 15, the partial sums h_m eventually get arbitrarily large. This important example highlights the difficulty of using numerical methods to reliably investigate the convergence of series. To get definitive results, we need some additional theory as developed in the remainder of this chapter.

The next proposition says that the value $p = 1$ is a sort of dividing line between divergent and convergent p-series, as the recent examples have suggested.

PROPOSITION 3.18 (p-series). *The convergence / divergence behavior of p-series $\sum_{n=1}^{\infty} 1/n^p$ may be summarized as follows:*

(a) if $p > 1$, then $\sum_{n=1}^{\infty} 1/n^p$ converges;
(b) if $p \leq 1$, then $\sum_{n=1}^{\infty} 1/n^p$ diverges.

We postpone the proof until after we have developed a theoretical tool called *the integral test*. The basic idea will be to compare the p-series $\sum_{n=1}^{\infty} 1/n^p$ with the improper integral $\int_1^{\infty} \frac{1}{x^p} dx$, so we begin by briefly reviewing the idea of improper integration from MATH 155.

Suppose that $f\colon (0, \infty) \to \mathbb{R}$ is a real function defined on the positive real axis. Because it is the case that will be important for us, we assume that the function f is nonnegative, so that its graph lies entirely above the x-axis. We also assume that the function f is continuous, so that its integral over any finite interval $[a, b]$ exists. But we would like to compute the area under the graph of f over the *infinite* interval $[1, \infty)$. This is a region of infinite extent, so you might initially guess that its area must also be infinite. But in fact the area is sometimes finite, as revealed by the following procedure: fix (for the moment) a large upper bound $t > 0$, and consider the finite integral

$$A(t) = \int_1^t f(x) dx.$$

The function $A(t)$ gives the area under f over the interval $[1, t]$, as shown in Figure 3.3.

Now take the limit of $A(t)$ as $t \to \infty$, and define the *improper integral*

$$\int_1^{\infty} f(x) dx = \lim_{t \to \infty} A(t).$$

If $\lim_{t\to\infty} A(t) = A$ is finite, then we say that the improper integral converges. On the other hand, if $\lim_{t\to\infty} A(t) = +\infty$, then we say that the improper integral diverges.

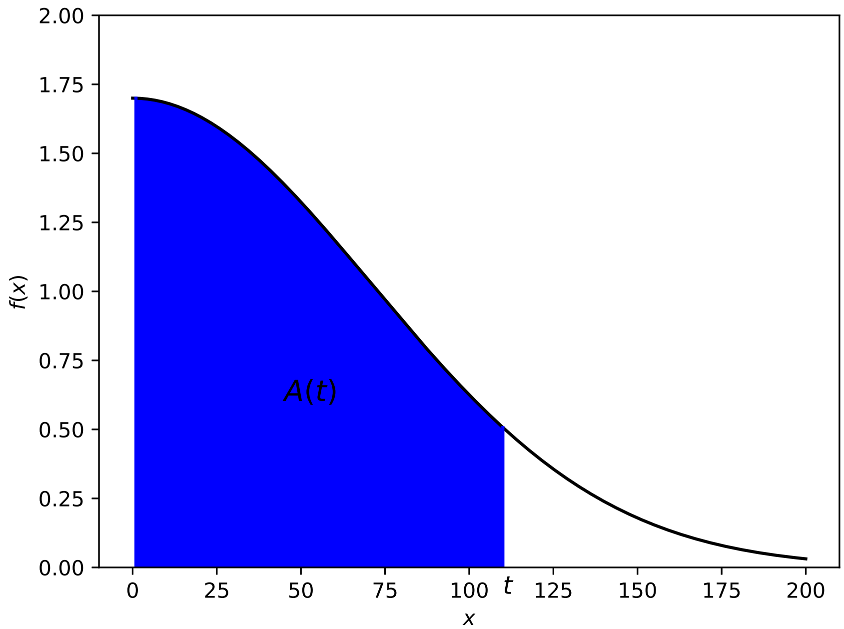


FIGURE 3.3. The finite areas $A(t)$ account for more and more of the total area under the graph of $f(x)$ as $t \to \infty$.

The justification for this procedural definition is similar to the justification for infinite series given on page 113: if the areas $A(t)$ approach a finite limit A as t gets large, it makes sense to think of A as the total area of the infinite region, since the intervals $[1, t]$ account for more and more of the interval $[1, \infty)$ as t gets large. On the other hand, since the total area must always be somewhat greater than any particular area $A(t)$, if the areas $A(t)$ run off to $+\infty$, we should conclude that the total area is also infinite.

As discussed at the end of Section 2.5, many real sequences (a_n) arise from real functions $f(x)$ by setting $a_n = f(n)$. For instance, the p-sequence $(a_n) = (1/n^p)$ is related to the function $f(x) = 1/x^p$ in this way. We now state a result known as *the integral test*, because it uses improper integrals to provide a criterion for the convergence or divergence of some real series $\sum_{n=1}^{\infty} a_n$ arising from nonnegative continuous functions $f(x)$.

PROPOSITION 3.19 (The Integral Test). *Suppose that $f(x)$ is a continuous, nonnegative, and decreasing function defined on the positive x-axis, and consider the associated sequence $(a_n) = (f(n))$. Then:*

(a) the series $\sum_{n=1}^{\infty} a_n$ converges if the improper integral $\int_1^{\infty} f(x)dx$ converges;

(b) the series $\sum_{n=1}^{\infty} a_n$ diverges if the improper integral $\int_1^{\infty} f(x)dx$ diverges.

REMARK 3.20. Be careful to check the hypotheses carefully when using the integral test: the real function $f\colon (0,\infty) \to \mathbb{R}$ must satisfy all three of the stated conditions: continuous, nonnegative, and decreasing.

EXERCISE 3.4. After reading the proof of the integral test, write a paragraph explaining where each of the hypotheses on the function f are used in the proof: (1) continuous, (2) nonnegative, (3) decreasing.

PROOF. We have two things to prove: (a) If the improper integral converges, then so does the series; (b) if the improper integral diverges, then so does the series. We begin with (a).

Assume that the improper integral converges to a finite area A, so $\int_1^{\infty} f(x)dx = A < +\infty$. We wish to show that the sequence of partial sums (s_m) of the series $\sum_{n=1}^{\infty} a_n$ converges. For this we will use the monotone convergence theorem (Theorem 2.43), so we need to show that the sequence (s_m) is increasing and bounded. Increasing follows from the fact that the terms $a_n = f(n)$ are nonnegative:

$$s_{m+1} = s_m + a_{m+1} \geq s_m.$$

Figure 3.4 shows that each partial sum s_m is less than the upper bound $a_1 + A$. By the monotone convergence theorem, the sequence (s_m) converges.

For part (b), we start by assuming that the improper integral diverges, so that $\lim_{t\to\infty} A(t) = +\infty$. Figure 3.5 shows that the partial sum $s_m > A(m+1)$. Since the numbers $A(m+1)$ get arbitrarily large,

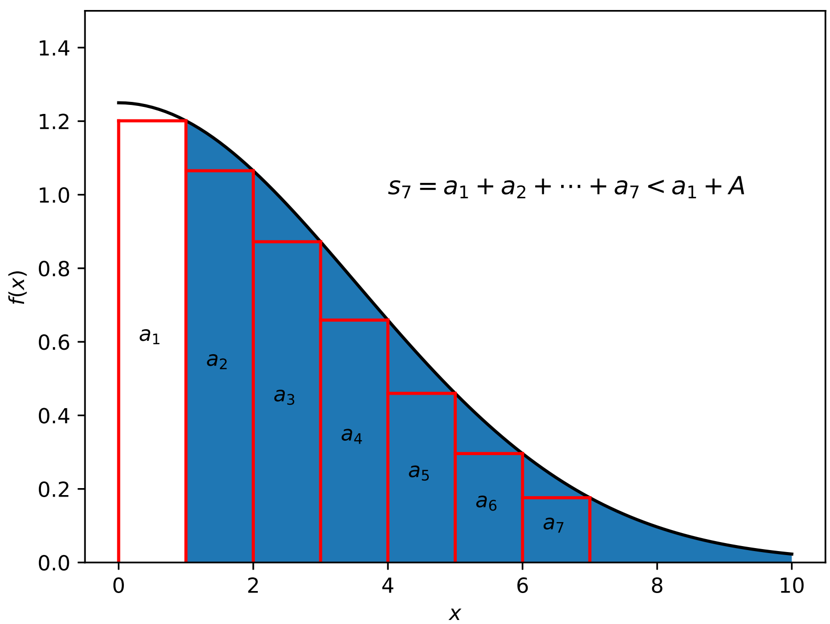


FIGURE 3.4. The blue region under the graph of f is the total area A, while the sum of the areas of the 7 red rectangles is the partial sum s_7. In general, for every index m, we have $s_m < a_1 + A$.

it follows that the partial sums s_m also get arbitrarily large. That is, the sequence (s_m) diverges to $+\infty$. □

The problems at the end of the chapter ask you to use the integral test to determine the convergence / divergence of various series. Our main concern is to determine the behavior of p-series, so we now illustrate the use of the integral test by providing a proof of Proposition 3.18.

PROOF OF PROPOSITION 3.18. First of all, note that we have already established (using the divergence test in Example 3.9), that $\sum_{n=1}^{\infty} 1/n^p$ diverges for $p \leq 0$.

So we assume that the exponent $p > 0$; in this case the function $f(x) = 1/x^p$ satisfies the hypotheses of the integral test: it is continuous, nonnegative, and decreasing. We begin by investigating the

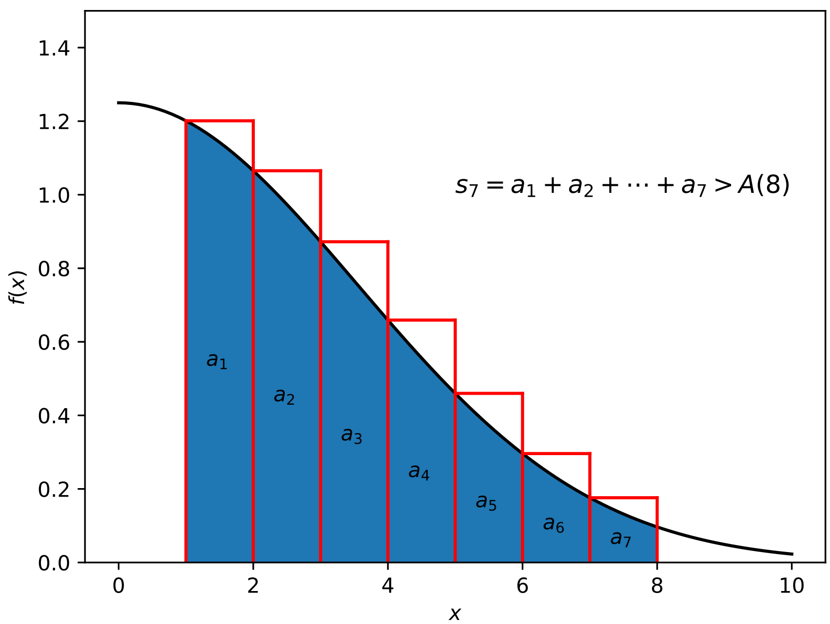


FIGURE 3.5. The blue region under the graph of f is the finite area $A(8)$, while the sum of the areas of the 7 red rectangles is the partial sum s_7. In general, for every index m, we have $s_m > A(m+1)$.

improper integral. First assume $p \neq 1$. Then

$$\begin{aligned}\int_1^\infty \frac{1}{x^p}dx = \lim_{t\to\infty}\int_1^t \frac{1}{x^p}dx &= \lim_{t\to\infty} \frac{x^{1-p}}{1-p}\bigg|_1^t \\ &= \lim_{t\to\infty}\frac{t^{1-p}-1}{1-p} = \begin{cases} +\infty & 0<p<1 \\ \frac{1}{p-1} & p>1. \end{cases}\end{aligned}$$

By the integral test, it follows that $\sum_{n=1}^\infty 1/n^p$ diverges for $p < 1$ and converges for $p > 1$

It remains to study the case $p = 1$, corresponding to the harmonic series:

$$\int_1^\infty \frac{1}{x}dx = \lim_{t\to\infty}\int_1^t \frac{1}{x}dx = \lim_{t\to\infty}\ln(x)\bigg|_1^t = \lim_{t\to\infty}(\ln(t)-\ln(1)) = +\infty.$$

Thus, the harmonic series $\sum_{n=1}^\infty 1/n$ diverges. □

The integral test only tells us whether the series converges, not the value of the sum. However, it does provide bounds for the sum. Figures 3.4 and 3.5 show that, in the case of convergence, we have

$$A \leq \sum_{n=1}^{\infty} a_n \leq a_1 + A,$$

where $A = \int_{n=1}^{\infty} f(x)dx$ is the finite value of the improper integral.

EXAMPLE 3.21. Consider the p-series $\sum_{n=1}^{\infty} 1/n^2$ studied earlier in Example 3.15. Then $a_1 = 1$ and

$$A = \int_1^{\infty} \frac{1}{x^2} dx = 1,$$

so we have

$$1 \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \leq 1 + 1 = 2,$$

which agrees with our earlier numerical estimate that $\sum_{n=1}^{\infty} \frac{1}{n^2} \approx 1.64$.

In the mid-18th century the Swiss mathematician Leonard Euler proved the amazing result that the exact value for this sum is

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} = 1.64493406684823\ldots$$

(This result goes by the name of the *Basel problem*). In fact, this is just the first level of Euler's discovery: for every even positive integer $p = 2k$, there is an explicitly known rational number r_k such that

$$\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = r_k \pi^{2k}.$$

For instance:

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}, \qquad \sum_{n=1}^{\infty} \frac{1}{n^6} = \frac{\pi^6}{945}, \quad \ldots, \quad \sum_{n=1}^{\infty} \frac{1}{n^{12}} = \frac{691\pi^{12}}{638512875}, \quad \ldots$$

Euler made his incredible discoveries while studying the real zeta function $\zeta(p)$ introduced in Section 3.1:

$$\zeta(p) = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^p}.$$

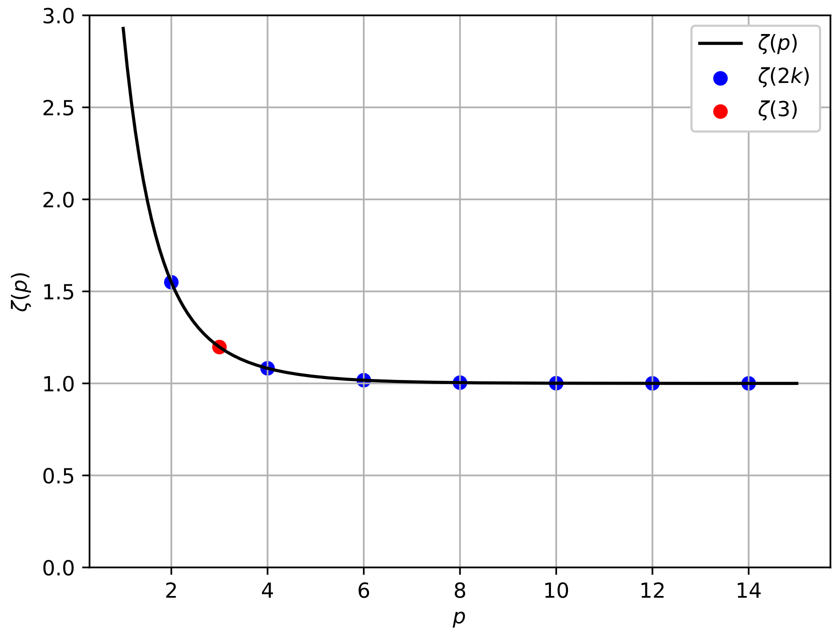


FIGURE 3.6. Graph of the real zeta function $\zeta(p)$. Even zeta values are shown as blue dots, and $\zeta(3)$ is shown as a red dot. This curve is what you would obtain by slicing the plot of the magnitude $|\zeta(z)|$ shown in Figure 3.1 with the vertical plane $y = 0$ lying above the real axis. The vertical asymptote at $p = 1$ indicates the divergence of the harmonic series.

We are now in a better position to appreciate this function. The defining formula on the right hand side is the p-series that we have been studying in this section. But now we are thinking of the real exponent p as a variable: as the value of p changes, so does the corresponding sum of the p-series $\sum_{n=1}^{\infty} 1/n^p$, and we call the resulting function $\zeta(p)$. The main result of this section (Proposition 3.18) says that the domain of $\zeta(p)$ is the infinite interval $(1, \infty)$. Figure 3.6 shows the graph of the real zeta function.

The fact that the so-called *even zeta values* $\zeta(2k) = \sum_{n=1}^{\infty} 1/n^{2k}$ can be explicitly written in terms of the familar constant π is amazing, and most zeta values can't be expressed in any such simple, finite form. For example, consider $\zeta(3) = \sum_{n=1}^{\infty} 1/n^3$. In this case the corresponding improper integral yields

$$A = \int_1^{\infty} \frac{dx}{x^3} = \frac{1}{2},$$

and so $0.5 \leq \zeta(3) \leq 1.5$, using the bounds coming from the integral test. Of course, we can get a better idea of the value $\zeta(3)$ by computing a partial sum, say for $m = 100$:

$$s_{100} = \sum_{n=1}^{100} \frac{1}{n^3} \approx 1.2020074.$$

But how do we know how accurate this partial sum is? Well, the error is given by the sum of all the terms that we did not include in the partial sum:

$$\text{error} = \zeta(3) - s_{100} = \sum_{n=1}^{\infty} \frac{1}{n^3} - \sum_{n=1}^{100} \frac{1}{n^3} = \sum_{n=101}^{\infty} \frac{1}{n^3},$$

so we would like to get an upper bound for the *tail* of the series, i.e. the sum of all terms starting with the 101st. But a version of Figure 3.4 with $m+1$ for the lower limit of integration instead of 1 would yield the inequality:

$$\sum_{n=m+1}^{\infty} \frac{1}{n^3} \leq \frac{1}{(m+1)^3} + \int_{m+1}^{\infty} \frac{dx}{x^3} = \frac{1}{(m+1)^3} + \frac{1}{2(m+1)^2}.$$

For the particular case $m = 100$, we find that (to 9 decimal places of precision):

$$\text{error} = \sum_{n=101}^{\infty} \frac{1}{n^3} \leq \frac{1}{101^3} + \frac{1}{2 \cdot 101^2} \approx 0.000049985$$

It follows that the partial sum approximation s_{100} is correct to at least 4 decimal places:

$$\zeta(3) \approx 1.2020$$

The number $\zeta(3)$ is known as *Apéry's constant*, after the French mathematician Roger Apéry who in 1978 proved that $\zeta(3)$ is an irrational number. Although it is known that there must be infinitely many other irrational *odd zeta values* $\zeta(2k+1)$, no other particular odd zeta value has yet been proven to be irrational.

Key points for Section 3.3:

- p-series convergence / divergence (Proposition 3.18)
- Integral Test (Proposition 3.19)

3.4. Comparison (ℝ)

This section takes place entirely in the context of the real numbers $\mathbb{R}$.

We have now established the convergence / divergence of two important families of series with positive terms:

- (p-series)
$$\sum_{n=1}^{\infty} \frac{1}{n^p} = \begin{cases} \zeta(p) & p > 1 \\ +\infty & p \leq 1 \end{cases}$$
- (positive geometric series) for $a, r > 0$ fixed real numbers,
$$\sum_{n=0}^{\infty} ar^n = \begin{cases} \frac{a}{1-r} & 0 < r < 1 \\ +\infty & r \geq 1. \end{cases}$$

The next result allows us to determine the convergence / divergence of other nonnegative series by comparing them to nonnegative series that we already understand.

Proposition 3.22 (The Comparison Test). *Suppose that* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *are two series with eventually nonnegative terms* a_n *and* b_n*. Furthermore, suppose that* $0 \leq a_n \leq b_n$ *for eventually all indices* n*. Then*

(1) if $\sum_{n=1}^{\infty} b_n$ *converges, so does* $\sum_{n=1}^{\infty} a_n$
(2) if $\sum_{n=1}^{\infty} a_n$ *diverges, so does* $\sum_{n=1}^{\infty} b_n$*.*

Moreover, in the convergent case (1), if all *terms are nonnegative and satisfy* $0 \leq a_n \leq b_n$*, then we have* $\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n$*.*

Remark 3.23. The word *eventually* is shorthand for more precise statements:

(a) A sequence (a_n) is *eventually* nonnegative if there exists an index N such that $a_n \geq 0$ for all $n \geq N$;
(b) the inequalities $0 \leq a_n \leq b_n$ hold for *eventually* all n if there exists an index N such that $0 \leq a_n \leq b_n$ for all $n \geq N$.

We will often use the word *eventually* to simplify the formulation of statements in this way. The point is that removing finitely many terms from a sequence (a_n) cannot affect whether or not the series $\sum_{n=0}^{\infty} a_n$ converges (although it does change the value of the sum). Thus, many hypotheses for convergence tests are stated as eventual properties of the terms a_n.

Remark 3.24. Parts (1) and (2) are are logically equivalent. Indeed, (1) has the form [B implies A] where B is the assertion that $\sum_{n=1}^{\infty} b_n$ converges and A is the assertion that $\sum_{n=1}^{\infty} a_n$ converges. In these terms, (2) is the statement [not A implies not B] which is the *contrapositive* of [B implies A]. And as a matter of logic, statements and their contrapositives are equivalent.

Explicitly, if [B implies A] is true and [not A] is also true, then it is impossible for B to be true, for then both A and [not A] would be true, a contradiction. Hence the statement [B implies A] entails its contrapositive [not A implies not B]. The same argument shows that the statement [not A implies not B] entails [B implies A].

Proof. By the previous remark, we only need to prove part (1), the convergent case. Moreover, since the question of series convergence or divergence depends only on the *eventual* behavior of the terms, we may assume that *all* terms are nonnegative and that $0 \leq a_n \leq b_n$ for all n. Let (s_m) denote the sequence of partial sums of $\sum_{n=1}^{\infty} a_n$, and (t_m) the sequence of partial sums of $\sum_{n=1}^{\infty} b_n$.

For (1), we are assuming that the sequence (t_m) converges to a finite sum t, and we must show that the sequence (s_m) converges. We will use the monotone convergence theorem. The sequence (s_m) is increasing, since the terms a_n are nonnegative:

$$s_{m+1} = s_m + a_{m+1} \geq s_m.$$

For the same reason, the sequence of partial sums (t_m) is increasing, which means that the sum t is an upper bound for that sequence: $t_m \leq t$ for all indices m. But then

$$s_m = a_1 + a_2 + \cdots a_m \leq b_1 + b_2 + \cdots + b_m = t_m \leq t,$$

which shows that t is an upper bound for the sequence (s_m). Hence, the sequence of partial sums (s_m) is increasing and bounded above, so convergent by the monotone convergence theorem. Morever, by EXERCISE 2.12 from Section 2.6, it follows that

$$\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} s_m \leq t = \sum_{n=1}^{\infty} b_n.$$

□

EXAMPLE 3.25. Consider the series

$$\sum_{n=1}^{\infty} \frac{n}{n^3 + n^2 + 1}.$$

The strategy here is to choose the p-series that most resembles this series, and then try to use the comparison test. If we pay attention only to the leading terms of the numerator and denominator, the ratio looks like $n/n^3 = 1/n^2$, and so we choose $p = 2$. The p-series $\sum_{n=1}^{\infty} 1/n^2$ is convergent, so we try to use part (1) of the comparison test, with $a_n = n/(n^3 + n^2 + 1)$ and $b_n = 1/n^2$. Note that

$$\frac{n}{n^3 + n^2 + 1} < \frac{n}{n^3} = \frac{1}{n^2},$$

so the required inequality does in fact hold. We conclude that the series converges.

EXAMPLE 3.26. Now consider the series

$$\sum_{n=1}^{\infty} \frac{n^2}{n^3 - n}.$$

Again, paying attention just to the leading terms of the numerator and denominator yields the ratio $n^2/n^3 = 1/n$. So we will compare with the divergent harmonic series $\sum_{n=1}^{\infty} 1/n$, using part (2) of the comparison test with $a_n = 1/n$ and $b_n = n^2/(n^3 - n)$. Since $n^3 - n < n^3$ for all n, we have

$$\frac{n^2}{n^3 - n} > \frac{n^2}{n^3} = \frac{1}{n},$$

which is the required inequality. So we conclude that the series diverges.

EXAMPLE 3.27. Now consider the series

$$\sum_{n=1}^{\infty} \frac{1}{n + \sqrt{n}}.$$

In this case, we might try $p = 1$, since when n is large $\sqrt{n}$ is much smaller than n. So we would like to compare to the divergent harmonic series, using part (2) of the comparison test with $a_n = 1/n$ and $b_n = 1/(n + \sqrt{n})$. Unfortunately, $1/n > 1/(n + \sqrt{n})$, so the inequality goes the wrong way! But note that $n + \sqrt{n} < 2n$, so $1/2n < 1/(n + \sqrt{n})$ as required. Moreover, the series $\sum_{n=1}^{\infty} \frac{1}{2n}$ also diverges (this follows either from the integral test or from Proposition 3.11). So, we conclude that $\sum_{n=1}^{\infty} \frac{1}{n+\sqrt{n}}$ diverges.

REMARK 3.28. The previous example shows that our first guess for a comparison series may not quite work, and it may be necessary to modify our initial guess in some way. In the optional Section 3.7, we present a more sophisticated comparison method called the *limit comparison test* that often eliminates the need for clever modifications of the initial guess.

The problems at the end of the chapter ask you to use the comparison test to determine the convergence / divergence of various series. Our main theoretical use for the comparison test is to develop two powerful tools that will help us establish the convergence of many complex series. We take up the first of these tools in the next section, where we introduce and study the notion of *absolute convergence* of a complex series. The second tool is called *the ratio test* and is the subject of Section 3.6.

Key points for Section 3.4:

- Comparison Test (Proposition 3.22)

3.5. Absolute and Conditional Convergence

The integral and comparison tests apply only to nonnegative real series. But we are interested more generally in series with complex terms. Fortunately, it turns out that we can often establish the convergence of a complex series by instead investigating the convergence of a related nonnegative real series, formed by taking the magnitude of each term. To explain this further, we introduce a definition.

DEFINITION 3.29. Suppose that $\sum_{n=1}^{\infty} c_n$ is a complex series, and consider the series obtained by replacing each term c_n by its magnitude $|c_n|$:

$$\sum_{n=1}^{\infty} |c_n|.$$

If this nonnegative real series converges, then we say that the original series $\sum_{n=1}^{\infty} c_n$ is *absolutely convergent.*

PROPOSITION 3.30. *If the complex series $\sum_{n=1}^{\infty} c_n$ is absolutely convergent, then it is convergent.*

PROOF. Suppose that $\sum_{n=1}^{\infty} |c_n|$ converges. We need to show that the original series $\sum_{n=1}^{\infty} c_n$ converges. We will do this by considering the real and imaginary parts of the sequence of terms (c_n); we will write $c_n = a_n + b_n i$.

We first do the special case where each of the terms $c_n = a_n$ is real. Under this assumption, note that we have the following inequalities:

$$0 \le a_n + |a_n| \le 2|a_n|.$$

Indeed, if $a_n \ge 0$, then $|a_n| = a_n$ and so $a_n + |a_n| = 2|a_n| \ge 0$. On the other hand, if $a_n < 0$ is negative, then $|a_n| = -a_n > 0$ is positive, and we have $a_n + |a_n| = a_n - a_n = 0 < 2|a_n|$. In either case, the stated inequalities hold.

Our assumption is that the series $\sum_{n=1}^{\infty} |a_n|$ converges, from which it follows that the constant multiple series $\sum_{n=1}^{\infty} 2|a_n|$ also converges (Proposition 3.11(c)). By the comparison test, we then conclude that

$\sum_{n=1}^{\infty}(a_n+|a_n|)$ converges. Using Proposition 3.11(b), we see our original series converges, as claimed:

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty}((a_n+|a_n|)-|a_n|) = \sum_{n=1}^{\infty}(a_n+|a_n|) - \sum_{n=1}^{\infty}|a_n|.$$

Now return to the general case where the terms $c_n = a_n + b_n i$ are complex. Then by Proposition 1.6, we have $|a_n| \le |c_n|$ and $|b_n| \le |c_n|$. Since we are assuming that $\sum_{n=1}^{\infty}|c_n|$ converges, the comparison test says that each of the series $\sum_{n=1}^{\infty}|a_n|$ and $\sum_{n=1}^{\infty}|b_n|$ are also convergent. By the special case above, we find that the real series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converge. Since these are the real and imaginary parts of our original complex series, it follows from Proposition 3.12 that $\sum_{n=1}^{\infty} c_n$ converges. □

Proposition 3.30 implies that, even though the integral and comparison tests apply only to nonnegative real series, these tests are actually able to establish the convergence of more general complex series, by showing that they are absolutely convergent. We illustrate with the following example.

Example 3.31. Consider the following complex series:

$$\sum_{n=1}^{\infty} \frac{\left(\frac{3}{5}+\frac{4}{5}i\right)^n}{n^2+in}.$$

Let's show that it is absolutely convergent. We start by computing the magnitude of the nth term:

$$\left|\frac{\left(\frac{3}{5}+\frac{4}{5}i\right)^n}{n^2+in}\right| = \frac{\left|\left(\frac{3}{5}+\frac{4}{5}i\right)\right|^n}{|n^2+in|} = \frac{1}{\sqrt{n^4+n^2}}.$$

We may compare the series of magnitudes $\sum_{n=1}^{\infty} 1/\sqrt{n^4+n^2}$ to the convergent p-series $\sum_{n=1}^{\infty} 1/n^2$, because

$$\frac{1}{\sqrt{n^4+n^2}} = \frac{1}{n\sqrt{n^2+1}} < \frac{1}{n^2}.$$

By the comparison test, the series of magnitudes $\sum_{n=1}^{\infty} 1/\sqrt{n^4+n^2}$ converges, which means that the original complex series is absolutely convergent. By Proposition 3.30, the original complex series converges.

REMARK 3.32. Note that Proposition 3.30 only tells us that the series converges, without telling us much about the sum. In particular, the magnitude of the sum of the series will not be equal to the sum of the series of magnitudes. For an explicit example, consider the convergent geometric series $\sum_{n=0}^{\infty}(i/2)^n$. By Example 3.10, we have

$$\sum_{n=0}^{\infty}\left(\frac{i}{2}\right)^n = \frac{1}{1-\frac{i}{2}} = \frac{2}{2-i} = \frac{2}{5}(2+i).$$

The magnitude of this sum is $2/\sqrt{5}$.

On the other hand, the sequence of magnitudes is another convergent geometric series:

$$\sum_{n=0}^{\infty}\left|\frac{i}{2}\right|^n = \sum_{n=0}^{\infty}\frac{1}{2^n} = \frac{1}{1-\frac{1}{2}} = 2.$$

Absolute convergence is a special type of convergence, and there are convergent series that are not absolutely convergent. We give these a name in the following definition.

DEFINITION 3.33. If a complex series converges but is not absolutely convergent, we say that it is *conditionally convergent*.

REMARK 3.34. Conditional convergence of a series $\sum_{n=1}^{\infty} c_n$ is unstable, in the sense that it depends on a tug-of-war between the arguments of the terms c_n to achieve convergence. If we eliminate the tug-of-war by using only the magnitudes $|c_n|$, then the series diverges. In the optional Section 3.8, we explain a surprising aspect of the instability of conditionally convergent series. In particular, we will explain how it is possible to reorder the terms c_n so as to achieve different sums! Such strange behavior does not occur for absolutely convergent series, and in this sense absolute convergence is "better" than conditional convergence.

EXAMPLE 3.35. (alternating harmonic series) The most famous example of a conditionally convergent series is the *alternating harmonic series*

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

The series of magnitudes is the divergent harmonic series $\sum_{n=1}^{\infty} 1/n$, so the alternating harmonic series is not absolutely convergent. But our next result (the *alternating series test*) says that it does converge, and hence provides an example of a conditionally convergent series. Before stating this result, we need a definition.

DEFINITION 3.36. A real series $\sum_{n=1}^{\infty} a_n$ is *alternating* if the terms alternate in sign. That is, the series is alternating if there is a nonnegative sequence (b_n) such that either

$$a_n = (-1)^{n+1} b_n \qquad \text{for } n \geq 1,$$

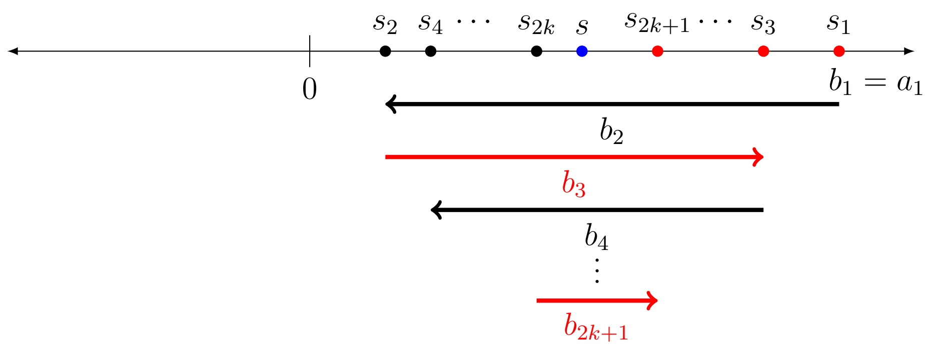
or

$$a_n = (-1)^n b_n \qquad \text{for } n \geq 1.$$

(The two cases allow for the first term a_1 to be either positive or negative.)

PROPOSITION 3.37 (Alternating Series Test). *Suppose that* $\sum_{n=1}^{\infty} a_n$ *is an alternating series with* $\lim_{n\to\infty} a_n = 0$. *Moreover, suppose that the sequence of magnitudes* $(b_n) = (|a_n|)$ *is eventually decreasing:* $b_n \geq b_{n+1}$ *for eventually all* n. *Then the series converges.*

PROOF. We will prove the case where a_1 is positive, the other case being similar. The following picture will be useful as we construct the proof:



Consider the partial sums s_{2k} with even indices, and remember that the even terms a_{2k} are negative, the odd terms a_{2k+1} are positive, and the magnitudes $b_n = |a_n|$ are decreasing:

$$\begin{aligned} s_2 &= a_1 + a_2 = b_1 - b_2 \geq 0 \\ s_4 &= a_1 + a_2 + a_3 + a_4 = s_2 + (b_3 - b_4) \geq s_2 \\ s_6 &= s_4 + a_5 + a_6 = s_4 + (b_5 - b_6) \geq s_4 \\ &\vdots \end{aligned}$$

We see that the sequence (s_{2k}) of even partial sums is increasing. But these numbers are all bounded above by $b_1 = a_1$:

$$s_{2k} = b_1 - (b_2 - b_3) - \cdots - (b_{2(k-1)} - b_{2k-1}) - b_{2k} \leq b_1.$$

Thus, the sequence of even partial sums (s_{2k}) is increasing and bounded, so (by the monotone convergence theorem) it converges to a limit which we call s. But now consider the sequence of odd partial sums (s_{2k+1}):

$$s_{2k+1} = s_{2k} + a_{2k+1} = s_{2k} + b_{2k+1}.$$

We have just shown that the sequence (s_{2k}) converges to s. Moreover, we are assuming that the sequence of terms (a_n) converges to zero, i.e. that the terms a_n get arbitrarily small. But then if we only pay attention to the odd terms a_{2k+1}, we see that these are also getting arbitrarily small, which means that the sequence containing only the odd terms (a_{2k+1}) also converges to zero. An application of Proposition 2.20

then shows that

$$\lim_{k\to\infty} s_{2k+1} = \lim_{k\to\infty}(s_{2k} + a_{2k+1}) = \lim_{k\to\infty} s_{2k} + \lim_{k\to\infty} a_{2k+1} = s + 0 = s.$$

Since the two sequences (s_{2k}) and (s_{2k+1}) together comprise the entire sequence of partial sums, it follows that the full sequence of partial sums (s_m) converges to s, which means that the alternating series $\sum_{n=1}^{\infty} a_n$ converges. □

As with the integral and comparison tests, the alternating series test tells us only that the series converges, but does not explicitly identify the sum. But it turns out that the sums of many convergent alternating series are especially easy to estimate. We illustrate with the alternating harmonic series.

EXAMPLE 3.38. Denote the sum of the alternating harmonic series by s, so that

$$s = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}.$$

Each partial sum s_m provides an approximation for s, but how good are they? Well, the size of the error is given by the absolute value of the tail of the series:

$$|\text{error}(m)| = |s - s_m| = \left| \sum_{n=m+1}^{\infty} \frac{(-1)^{n+1}}{n} \right|.$$

But note that the sequence of magnitudes $(1/n)$ is decreasing. By the alternating nature of the sum, it follows that the size of the tail of the series is at most the size of its first term, $1/(m+1)$. So we see that for all $m \geq 1$

$$|\text{error}(m)| < 1/(m+1).$$

In general: *for convergent alternating series with decreasing terms, the error in the mth partial sum approximation is bounded by the size of the $(m+1)$st term.*

For example, suppose we wish to determine the value of the alternating harmonic series $s = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$, correct to 4 decimal places. So we take $m = 10^5$ and look at the corresponding partial sum, keeping

6 decimal places of precision:

$$s_{10^5} \approx 0.693152.$$

Since the error in this approximation is less than $1/(10^5+1) < 0.00001$, we see that the first four decimal places of s are correctly given by $s \approx 0.6931$

We end this section by mentioning the tantalizing fact that

$$\ln(2) = 0.693147\ldots.$$

Could it be that the alternating harmonic series converges to the natural logarithm of 2?

Key points for Section 3.5:

- Absolute convergence (Definition 3.29)
- Absolute convergence implies convergence (Proposition 3.30)
- Conditional convergence (Definition 3.33)
- Alternating series and test (Definition 3.36 and Proposition 3.37)
- Conditional convergence of alternating harmonic series (Example 3.35)

3.6. The Ratio Test

In this section, we establish a powerful test for absolute convergence; the key step of the proof involves comparison with a convergent geometric series.

Proposition 3.39 (The Ratio Test). *Suppose that $\sum_{n=1}^{\infty} c_n$ is a complex series with nonzero terms c_n. Investigate the consecutive ratios of magnitudes $|c_{n+1}|/|c_n|$:*

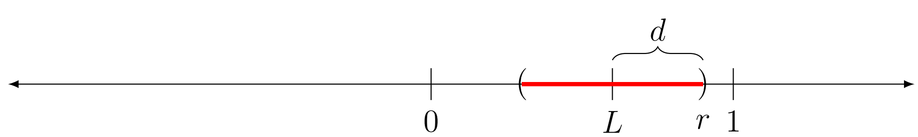
(1) If $\lim_{n\to\infty} |c_{n+1}|/|c_n| = L < 1$, then the series $\sum_{n=1}^{\infty} c_n$ is absolutely convergent;

(2) If $\lim_{n\to\infty} |c_{n+1}|/|c_n| = L > 1$ *or if* $\lim_{n\to\infty} |c_{n+1}|/|c_n| = +\infty$, *then the series* $\sum_{n=1}^{\infty} c_n$ *diverges.*

Remark 3.40. If $\lim_{n\to\infty} |c_{n+1}|/|c_n| = 1$, then we can't draw any definite conclusion from the ratio test: the series may be convergent or divergent. For example:

- the p-series $\sum_{n=1}^{\infty} 1/n^2$ converges, but
$$\lim_{n\to\infty} \frac{n^2}{(n+1)^2} = \lim_{n\to\infty} \frac{1}{1+\frac{2}{n}+\frac{1}{n^2}} = 1.$$
- the harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges, but
$$\lim_{n\to\infty} \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = 1.$$

Proof of Theorem 3.39. To prove (1), we assume that the limit $\lim_{n\to\infty} |c_{n+1}|/|c_n| = L < 1$. Our goal is to compare the series of magnitudes $\sum_{n=0}^{\infty} |c_n|$ to a convergent geometric series. We begin by choosing a positive real number r with $L < r < 1$; this will be the common ratio of our geometric series. Consider the interval of radius $d = r - L > 0$ centered at L:



Since the sequence $(|c_{n+1}|/|c_n|)$ converges to L, it follows that there is an index N such that for all $n \geq N$, then ratios $|c_{n+1}|/|c_n|$ are contained in this interval. In particular, for $n \geq N$ we have
$$\frac{|c_{n+1}|}{|c_n|} \leq r < 1.$$

Multiplying by $|c_n|$ yields the relation $|c_{n+1}| \leq r|c_n|$ for $n \geq N$. So we have

$$\begin{aligned}
|c_{N+1}| &\leq r|c_N| \\
|c_{N+2}| &\leq r|c_{N+1}| \leq r^2|c_N| \\
|c_{N+3}| &\leq r|c_{N+2}| \leq r^3|c_N| \\
&\vdots \\
|c_n| &\leq r^{n-N}|c_N| \\
&\vdots
\end{aligned}$$

Now watch the indices carefully to recognize the convergent geometric series:

$$\begin{aligned}
\sum_{n=N}^{\infty} r^{n-N}|c_N| &= |c_N| + |c_N|r + |c_N|r^2 + |c_N|r^3 + \cdots \\
&= \sum_{k=0}^{\infty} |c_N|r^k.
\end{aligned}$$

Hence, by the comparison test we find that the series $\sum_{n=N}^{\infty} |c_n|$ converges. But this only differs from the full series of magnitudes $\sum_{n=1}^{\infty} |c_n|$ by finitely many terms, so it follows that $\sum_{n=1}^{\infty} |c_n|$ converges, and hence the original series $\sum_{n=1}^{\infty} c_n$ is absolutely convergent.

For (2), suppose that the limit $L > 1$ or $L = +\infty$. A similar argument to that given for (1) shows that there is an index N and a number $r > 1$ such that for $n \geq N$ we have $|c_{n+1}| \geq r|c_n|$. Then it follows as before that $|c_n| \geq r^{n-N}|c_N| \geq |c_N| > 0$ for $n \geq N$. Thus, the sequence of terms (c_n) does not converge to zero, so the series $\sum_{n=1}^{\infty} c_n$ diverges by the divergence test. □

Exercise 3.5. Fill in the details of the proof of part (2) of the ratio test, Proposition 3.39.

Example 3.41. Consider the positive series

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \frac{2^n}{n^2}.$$

To apply the ratio test, we begin by investigating the limit of the ratio of consecutive terms:

$$\begin{aligned}\lim_{n\to\infty}\frac{a_{n+1}}{a_n} &= \lim_{n\to\infty}\frac{2^{n+1}}{(n+1)^2}\frac{n^2}{2^n}\\ &= \lim_{n\to\infty}\frac{2n^2}{n^2+2n+1}\\ &= \lim_{n\to\infty}\frac{2}{1+\frac{2}{n}+\frac{1}{n^2}}\\ &= 2.\end{aligned}$$

Since the limit is greater than 1, the ratio test says that the series diverges.

EXAMPLE 3.42. Now consider the complex series

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty}\frac{(2i)^n}{3^n\sqrt{n}}.$$

We apply the ratio test:

$$\begin{aligned}\lim_{n\to\infty}\frac{|c_{n+1}|}{|c_n|} &= \lim_{n\to\infty}\frac{2^{n+1}}{3^{n+1}\sqrt{n+1}}\frac{3^n\sqrt{n}}{2^n}\\ &= \lim_{n\to\infty}\frac{2}{3}\sqrt{\frac{n}{n+1}}\\ &= \frac{2}{3}\lim_{n\to\infty}\sqrt{\frac{1}{1+\frac{1}{n}}}\\ &= \frac{2}{3}.\end{aligned}$$

Since the limit is less than 1, the ratio test says that the series is absolutely convergent.

EXAMPLE 3.43. The ratio test tends to work especially well on series whose terms c_n involve n as an exponent or in a factorial. For instance: fix a complex number c and consider the series

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty}\frac{c^n}{n!}.$$

To apply the ratio test, we compute the ratio of consecutive magnitudes:

$$\frac{|c_{n+1}|}{|c_n|} = \frac{|c|^{n+1}}{(n+1)!}\frac{n!}{|c|^n} = \frac{|c|}{n+1}.$$

Now take the limit, remembering that c is a fixed complex number:

$$\lim_{n\to\infty} \frac{|c|}{n+1} = |c| \lim_{n\to\infty} \frac{1}{n+1} = |c| \cdot 0 = 0 < 1.$$

By the ratio test, it follows that the series is absolutely convergent for all complex c.

We record a consequence of this example as a lemma for future use. It says that the factorials $n!$ grow much faster than the terms of any geometric sequence (c^n). For more about *growth rates*, see the optional Section 3.7.

Lemma 3.44. *For any complex number c, the sequence $(c^n/n!)$ converges to zero.*

Proof. This follows from the divergence test: since the series $\sum_{n=0}^{\infty} \frac{c^n}{n!}$ converges, the sequence of terms $(c^n/n!)$ must converge to zero. □

Remark 3.45. If we think of z as a complex variable, then the previous example defines a complex function $h\colon \mathbb{C} \to \mathbb{C}$, given by the series formula

$$h(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4 + \cdots$$

This is an important special case of a *power series*, the topic of the next chapter.

> Key points for Section 3.6:
>
> - The ratio test (Proposition 3.39)

3.7. Optional: Limit Comparison, Growth and Decay ($\mathbb{R}$)

This section takes place entirely in the context of the real numbers $\mathbb{R}$*.*

In Example 3.27 of Section 3.4, we considered the positive series

$$\sum_{n=1}^{\infty} \frac{1}{n+\sqrt{n}}.$$

Motivated by the fact that n is much bigger than $\sqrt{n}$ when n is large, we chose to compare with the divergent harmonic series $\sum_{n=1}^{\infty} 1/n$. But our initial attempt was frustrated by the fact that the inequality between terms goes the wrong way to conclude divergence by the comparison test: $1/(n+\sqrt{n}) < 1/n$. Our initial intuition was correct, however, and we succeeded by comparing to the divergent series $\sum_{n=1}^{\infty} 1/2n$ instead.

At base, our initial diagnosis was that the series $\sum_{n=1}^{\infty} 1/(n+\sqrt{n})$ and $\sum_{n=1}^{\infty} 1/n$ behave the same way (i.e., diverge), because the two sequences of terms $(1/(n+\sqrt{n}))$ and $(1/n)$ converge to zero at similar rates. One way to measure the difference between these "rates of decay" is to investigate the ratios:

$$\lim_{n\to\infty} \frac{1/(n+\sqrt{n})}{1/n} = \lim_{n\to\infty} \frac{n}{n+\sqrt{n}} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{\sqrt{n}}} = 1.$$

The key point here is that the limit is a nonzero real constant, which we interpret as saying that the two rates of decay toward zero are similar. Observe that a limit of zero for these ratios would instead mean that the numerators were getting small significantly *faster* than the denominators:

$$\lim_{n\to\infty} \frac{1/n^2}{1/n} = \lim_{n\to\infty} \frac{n}{n^2} = \lim_{n\to\infty} \frac{1}{n} = 0.$$

At the other extreme, if the ratios diverge to $+\infty$, then the numerators are getting small significantly *slower* than the denominators:

$$\lim_{n\to\infty} \frac{1/n}{1/n^2} = \lim_{n\to\infty} \frac{n^2}{n} = \lim_{n\to\infty} n = +\infty.$$

These ideas come together in the following strengthening of the comparison test.

PROPOSITION 3.46 (The Limit Comparison Test). *Suppose that* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *are two series with eventually nonnegative terms* a_n *and* b_n. *Furthermore, suppose that* $b_n > 0$ *for eventually all* n. *Consider the ratios* a_n/b_n, *and suppose that*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L \qquad \text{with } 0 < L < +\infty.$$

Then the series $\sum_{n=1}^{\infty} a_n$ *converges if and only if the series* $\sum_{n=1}^{\infty} b_n$ *converges.*

PROOF. The basic idea is to reduce this result to the direct comparison test, Proposition 3.22. For this, we need to transform our hypothesis about the limit of the ratios into some inequalities between the terms a_n and b_n of the two series. As usual, since convergence depends only on the eventual behavior of the terms, we may assume that all terms are nonnegative, with $b_n > 0$ for all n.

Because L is a positive real number, we may choose a smaller positive real number $0 < d < L$. We then play (and win!) the convergence game using the distance d: there exists an index N such that for $n \geq N$ we have

$$\left|\frac{a_n}{b_n} - L\right| < d.$$

This single inequality for the absolute value is really a chain of two inequalities:

$$-d < \frac{a_n}{b_n} - L < d.$$

Now do a bit of rearrangment, and recall that $L - d > 0$:

$$0 < (L-d)b_n < a_n < (L+d)b_n \qquad \text{for all } n \geq N.$$

These are the inequalities that we are looking for, and we now use the comparison test.

First assume that $\sum_{n=1}^{\infty} b_n$ converges. Then, $\sum_{n=N}^{\infty} b_n$ also converges, since we have simply omitted finitely many terms at the beginning. Multiplying by the constant $L + d$ then yields the convergent series $\sum_{n=N}^{\infty}(L+d)b_n$. By the comparison test, it follows that $\sum_{n=N}^{\infty} a_n$ converges, and hence (after adding finitely many initial terms) that the full series $\sum_{n=1}^{\infty} a_n$ also converges.

Now suppose that $\sum_{n=1}^{\infty} a_n$ converges. The next exercise asks you to show that $\sum_{n=1}^{\infty} b_n$ must also converge. □

EXERCISE 3.6. Finish the proof of the limit comparison test (Proposition 3.46) by showing that the convergence of $\sum_{n=1}^{\infty} a_n$ implies the convergence of $\sum_{n=1}^{\infty} b_n$.

EXAMPLE 3.47. Consider the positive series

$$\sum_{n=1}^{\infty} \frac{100n^2 + 2n + 17}{n^4 + n + 1}.$$

Focusing just on the leading exponents of the numerator and denominator yields the convergent p-series $\sum_{n=1}^{\infty} \frac{n^2}{n^4} = \sum_{n=1}^{\infty} \frac{1}{n^2}$. We use limit comparison:

$$\begin{aligned} \lim_{n\to\infty} \frac{(100n^2 + 2n + 17)/(n^4 + n + 1)}{1/n^2} &= \lim_{n\to\infty} \frac{n^2(100n^2 + 2n + 17)}{n^4 + n + 1} \\ &= \lim_{n\to\infty} \frac{100 + 2/n + 17/n^2}{1 + 1/n^3 + 1/n^4} \\ &= 100. \end{aligned}$$

Since the limit is a nonzero constant, we conclude that the original series converges.

The topic of *decay rates* introduced before the limit comparison test is an important subject in its own right, and there is a complementary notion of *growth rates*. In fact, decay and growth are two sides of the same coin: if I understand the rate at which a positive sequence (a_n) is decaying to zero, then I also understand the rate at which the reciprocals $(1/a_n)$ are growing toward $+\infty$. In the remainder of this section, we discuss the growth rates of various sequences, and in particular we distinguish between 4 types of growth: logarithmic, power, exponential, and factorial.

DEFINITION 3.48 (logarithmic growth). Consider the nonnegative sequence $(\ln(n))$ arising from the real function $f(x) = \ln(x)$, which diverges to $+\infty$. If (a_n) is any positive sequence such that

$$\lim_{n\to\infty} \frac{\ln(n)}{a_n} = L \qquad \text{for } 0 < L < +\infty,$$

then we say that (a_n) has *logarithmic growth*.

EXAMPLE 3.49. The sequence $(a_n) = (\ln(n^2))$ has logarithmic growth:

$$\lim_{n\to\infty} \frac{\ln(n)}{a_n} = \lim_{n\to\infty} \frac{\ln(n)}{\ln(n^2)} = \lim_{n\to\infty} \frac{\ln(n)}{2\ln(n)} = \lim_{n\to\infty} \frac{1}{2} = \frac{1}{2}.$$

DEFINITION 3.50 (power growth). Fix a positive exponent $r > 0$, and consider the nonnegative sequence (n^r) arising from the real function $f(x) = x^r$. This sequence diverges to $+\infty$ since $r > 0$. If (a_n) is any positive sequence such

$$\lim_{n\to\infty} \frac{n^r}{a_n} = L \qquad \text{for } 0 < L < +\infty,$$

then we say that (a_n) has *r-power growth*. We say that a positive sequence (a_n) has *power growth* if there exists $r > 0$ such that (a_n) has r-power growth.

EXAMPLE 3.51. The sequence $(a_n) = (n^3 + \sqrt{n})$ has 3-power growth (also called *cubic* growth):

$$\lim_{n\to\infty} \frac{n^3}{n^3 + \sqrt{n}} = \lim_{n\to\infty} \frac{1}{1 + \frac{1}{n^2\sqrt{n}}} = 1.$$

DEFINITION 3.52 (exponential growth). Fix a positive base $c > 1$, and consider the geometric sequence (c^n) arising from the real function $f(x) = c^x$, which diverges to $+\infty$. If (a_n) is any positive sequence such

$$\lim_{n\to\infty} \frac{c^n}{a_n} = L \qquad \text{for } 0 < L < +\infty,$$

then we say that (a_n) has *c-exponential growth*. We say that a positive sequence (a_n) has *exponential growth* if it has c-exponential growth for some $c > 1$.

EXAMPLE 3.53. The sequence (e^{n+100}) has e-exponential growth:

$$\lim_{n\to\infty} \frac{e^n}{e^{n+100}} = \lim_{n\to\infty} \frac{e^n}{e^n e^{100}} = \lim_{n\to\infty} \frac{1}{e^{100}} = e^{-100}.$$

DEFINITION 3.54 (factorial growth). Consider the nonnegative sequence $(n!)$, which diverges to $+\infty$. If (a_n) is any positive sequence such that

$$\lim_{n\to\infty} \frac{n!}{a_n} = L \qquad \text{for } 0 < L < +\infty,$$

then we say that (a_n) has *factorial growth*.

EXERCISE 3.7. Give an example of a sequence (a_n) that has factorial growth.

We wish to determine the relationship between these four different rates of growth. The key idea is presented in the following definition.

DEFINITION 3.55. Suppose that (a_n) and (b_n) are positive sequences, both diverging to $+\infty$. Then we say that *the growth of* (b_n) *dominates the growth of* (a_n) if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Intuitively, dominance means that the terms b_n grow much faster than the terms a_n.

EXAMPLE 3.56. The growth of (n^2) dominates the growth of (n):

$$\lim_{n\to\infty} \frac{n}{n^2} = \lim_{n\to\infty} \frac{1}{n} = 0.$$

More generally, if $0 < r < s$, then s-power growth dominates r-power growth:

$$\lim_{n\to\infty} \frac{n^r}{n^s} = \lim_{n\to\infty} \frac{1}{n^{s-r}} = 0.$$

EXERCISE 3.8. Suppose that $d > c > 1$, and show that d-exponential growth dominates c-exponential growth: $\lim_{n\to\infty}(c^n/d^n) = 0$.

Our goal is to show that (see Figure 3.7)

factorial growth dominates
exponential dominates
power dominates
logarithmic.

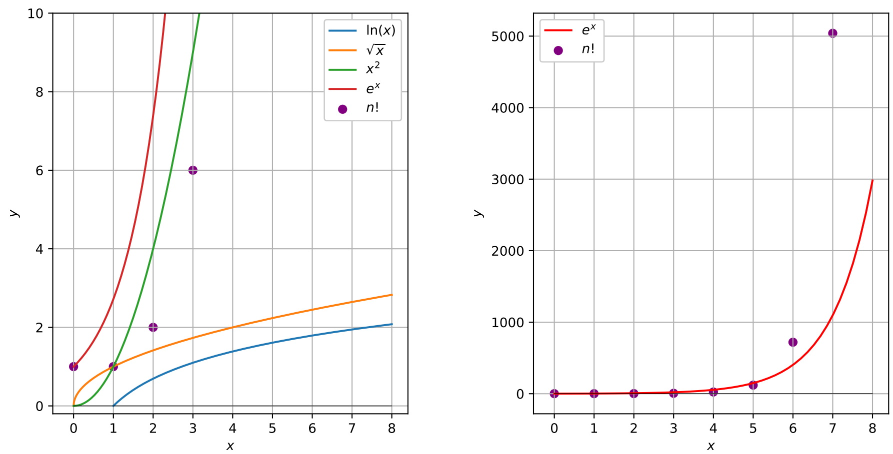


Figure 3.7. The left hand side shows plots indicating 5 different growth rates: logarithmic (blue), 1/2-power (orange), 2-power or quadratic (green), exponential (red), and factorial (purple dots). Dominance concerns the long term behavior of these plots, not the initial relationships: the square-root graph is initially above the quadratic graph, and the factorial sequence is initially below the exponential graph. The right hand side illustrates how factorial growth dominates exponential growth in the long-run.

We will begin at the bottom and work our way up.

Proposition 3.57 (power growth dominates logarithmic growth). *Let $r > 0$ be any positive real number. Then*

$$\lim_{n\to\infty} \frac{\ln(n)}{n^r} = 0.$$

As a result, if (a_n) has logarithmic growth and (b_n) has power growth, then the growth of (b_n) dominates the growth of (a_n).

Proof. The sequence $\big(\ln(n)/n^r\big)$ arises from $f(x) = \ln(x)/x^r$, so by Proposition 2.30, it suffices to prove that $\lim_{x\to\infty} f(x) = 0$. For this we use L'Hôpital's rule:

$$\lim_{x\to\infty} \frac{(\ln(x))'}{(x^r)'} = \lim_{x\to\infty} \frac{1/x}{rx^{r-1}} = \lim_{x\to\infty} \frac{1}{rx^r} = 0.$$

Thus, it follows that $\lim_{x\to\infty} f(x) = 0$, and so $\lim_{n\to\infty} \ln(n)/n^r = 0$.

Now suppose that (a_n) has logarithmic growth and (b_n) has power growth, with

$$\lim_{n\to\infty} \frac{\ln(n)}{a_n} = L_1 > 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{n^r}{b_n} = L_2 > 0.$$

Then

$$\begin{aligned} \lim_{n\to\infty} \frac{a_n}{b_n} &= \lim_{n\to\infty} \frac{a_n}{\ln(n)} \cdot \frac{\ln(n)}{n^r} \cdot \frac{n^r}{b_n} \\ &= \left(\lim_{n\to\infty} \frac{a_n}{\ln(n)}\right)\left(\lim_{n\to\infty} \frac{\ln(n)}{n^r}\right)\left(\lim_{n\to\infty} \frac{n^r}{b_n}\right) \\ &= \frac{1}{L_1} \cdot 0 \cdot L_2 \\ &= 0. \end{aligned}$$

This shows that the power growth of (b_n) dominates the logarithmic growth of (a_n). □

PROPOSITION 3.58 (exponential growth dominates power growth). *Let $r > 0$ be any positive real number, and $c > 1$ be a positive real base. Then*

$$\lim_{n\to\infty} \frac{n^r}{c^n} = 0.$$

As a result, if (a_n) has power growth and (b_n) has exponential growth, then the growth of (b_n) dominates the growth of (a_n).

PROOF. Consider the series $\sum_{n=1}^{\infty} \frac{n^r}{c^n}$ and apply the ratio test:

$$\lim_{n\to\infty} \frac{(n+1)^r}{c^{n+1}} \frac{c^n}{n^r} = \lim_{n\to\infty} \frac{1}{c}\left(\frac{n+1}{n}\right)^r = \frac{1}{c}\left(\lim_{n\to\infty}\left(1+\frac{1}{n}\right)\right)^r = \frac{1}{c} < 1,$$

since we are assuming that $c > 1$. By the ratio test, the series converges. But then the divergence test says that the sequence of terms (n^r/c^n) must converge to zero as claimed. □

EXERCISE 3.9. Finish the proof of Proposition 3.58 by modifying the end of the proof of Proposition 3.57.

PROPOSITION 3.59 (factorial growth dominates exponential growth). *Let $c > 1$ be a positive real base. Then*

$$\lim_{n\to\infty} \frac{c^n}{n!} = 0.$$

As a result, if (a_n) has exponential growth and (b_n) has factorial growth, then the growth of (b_n) dominates the growth of (a_n).

PROOF. The limit is Lemma 3.44. The final statement is proved as in Proposition 3.57 and EXERCISE 3.9. □

REMARK 3.60. These different growth rates are of fundamental importance for the study of *algorithms* in computer science. Roughly speaking, an algorithm is a precise sequence of steps for solving a mathematical problem for a given family of inputs. But algorithms are not all equally good, and in particular we are interested in efficient algorithms: those that will terminate in a reasonable amount of time. Of course, we expect the algorithm to take longer on bigger inputs than on smaller inputs, so what we care about is the growth rate of the sequence of running times (t_n) indexed by $n =$ size of input. Generally speaking, an algorithm is thought to be efficient if its sequence of running times (t_n) is dominated by power growth (these are called *polynomial time algorithms*). On the other hand, algorithms with (t_n) having exponential or factorial growth are considered extremely inefficient.

For example: consider the problem of determining whether a positive integer $m > 1$ is prime. A brute force algorithm consists of checking whether each of the numbers $1 < k < m$ is a divisor of m, and this takes approximately m steps (if we think of each trial division as one step). The relevant size of the input integer m is given roughly by $n = \log_2 m$, since integers are stored in a computer as strings of 0's and 1's using the binary number system. Thus, the sequence of running times (t_n) indexed by input size is given by

$$t_n \approx m = 2^{\log_2 m} = 2^n.$$

So the sequence of running times (t_n) of the brute force algorithm for testing primality has exponential growth, is not efficient, and has no practical value for the testing of large integers.

In 2002 there occurred a major development: the Indian computer scientists Manindra Agrawal, Neeraj Kayal, and Nitin Saxena published an efficient polynomial time algorithm for primality testing (now called the *AKS primality test*). They showed that the sequence of running times (t_n) for their algorithm is dominated by r-power growth for any $r > 12$, and variants of the algorithm have now reduced the value of r further, to around 6. You may be interested to learn that Kayal and Saxena were undergraduates when they discovered their algorithm!

Key points for Section 3.7:

- The limit comparison test (Proposition 3.46)
- Growth rates and dominance (Definitions 3.48–3.55)
- Factorial growth dominates exponential dominates power dominates logarithmic (Propositions 3.57–3.59 and Figure 3.7)

3.8. Optional: Rearrangements

In Section 3.5 we saw that the alternating harmonic series is conditionally convergent:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

Recall what this means: the series converges, but the corresponding sequence of magnitudes $\sum_{n=1}^{\infty} 1/n$ diverges. We also used the alternating nature of the series to provide an estimate for the value of the sum s, correct to 4 decimal places: $s \approx 0.6931$. In this section, we want to prove the following strange fact about the alternating harmonic series: simply by reordering the terms, we can make the sum any real number we choose.

Before stating and proving the result, we pause to emphasize how strange it is, based on our experience with *finite* summation. Indeed, if I have four numbers a, b, c, d, then the commutative and associative laws for addition imply that the order of summation doesn't matter:

$$a+b+c+d = a+c+d+b = d+c+b+a = \text{ etc.}$$

Clearly this extends to any finite collection of numbers: changing the order of a finite sum does not affect the result. But let's investigate what happens if we rearrange the alternating harmonic sequence as follows:

$$(a_\ell) = \left(1, -\frac{1}{2}, -\frac{1}{4}, \frac{1}{3}, -\frac{1}{6}, -\frac{1}{8}, \frac{1}{5}, -\frac{1}{10}, -\frac{1}{12}, \dots\right).$$

In this sequence, instead of alternating signs, we include two negative terms after each positive term. Here is the series:

$$\sum_{\ell=1}^{\infty} a_\ell = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \cdots$$

Let's look at some numerical values for the partial sums:

$$\begin{aligned} s_{2991} &\approx 0.346448 & s_{2996} &\approx 0.346699 \\ s_{2992} &\approx 0.346949 & s_{2997} &\approx 0.346448 \\ s_{2993} &\approx 0.346699 & s_{2998} &\approx 0.346949 \\ s_{2994} &\approx 0.346448 & s_{2999} &\approx 0.346699 \\ s_{2995} &\approx 0.346949 & s_{3000} &\approx 0.346449 \end{aligned}$$

It certainly appears that the partial sums are converging to a value close to 0.346, which is different than the sum $s \approx 0.6931$ of the alternating harmonic series.

PROPOSITION 3.61. *Let L be any real number. Then there exists a rearrangement (a_ℓ) of the alternating harmonic sequence $((-1)^{n+1}/n)$ such that $\sum_{\ell=1}^{\infty} a_\ell = L$. By a rearrangement, we mean that every term*

$(-1)^{n+1}/n$ appears exactly once in the sequence (a_ℓ), and there are no additional terms.

PROOF. We begin by observing that the series $\sum_{n=1}^{\infty} 1/2n$ diverges, since it is simply the divergent harmonic series multiplied by the constant $1/2$. But then the comparison test shows that the sum of just the positive terms of the alternating harmonic series also diverges:

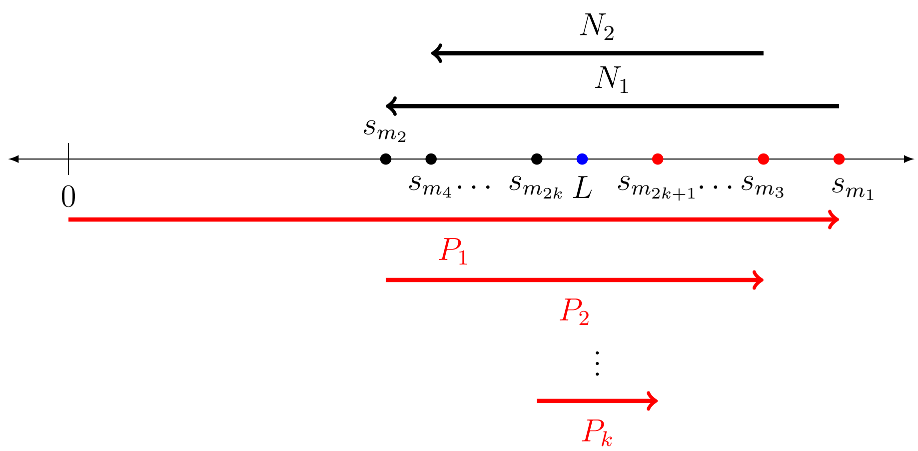
$$\text{Series } P: \qquad \sum_{n=1}^{\infty} \frac{1}{2n-1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots = +\infty.$$

Also, the sum of just the negative terms diverges:

$$\text{Series } N: \qquad \sum_{n=1}^{\infty} \frac{-1}{2n} = -\frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \cdots = -\infty.$$

Note that every term of the alternating harmonic sequence $((-1)^{n+1}/n)$ appears as a term in exactly one of the series P or N, and no other terms appear. So if we make use of every term from P and N exactly once in our new sequence (a_ℓ), then we will have a rearrangement of the alternating harmonic sequence.

We will assume that $L > 0$, although a similar argument works in the negative case. The basic idea is to use the positive terms from P until we get to the right of L, then make use of the negative terms from N until we get to the left of L, then use more terms from P until we overshoot L again, etc. The divergence statements above imply that no matter how many positive terms of the series P we have already used, the remaining terms still sum to $+\infty$. Similarly, no matter how many negative terms of the series N we have already used, the remaining terms still sum to $-\infty$. So we can continue this process forever, and at each stage we will be able to overshoot the target L. We build the sequence (a_ℓ) by listing the terms from P and N in the order we use them. As we will see, this construction leads to a sequence of partial sums (s_m) for the series $\sum_{\ell=1}^{\infty} a_\ell$ that converges to L. We describe this process a bit more explicitly below, by reference to the following picture:



(1) Let $m_1 \geq 1$ be the smallest index such that the finite sum $P_1 = \sum_{n=1}^{m_1} 1/(2n-1) > L$. We begin the sequence (a_ℓ) by using these terms:
$$(a_1, a_2, a_3, \ldots, a_{m_1}) = \left(1, \frac{1}{3}, \frac{1}{5}, \ldots, \frac{1}{2m_1 - 1}\right).$$
We have $s_{m_1} = P_1 > L$.

(2) Now let $k_1 \geq 1$ be the smallest index such that
$$P_1 + N_1 = P_1 - \sum_{n=1}^{k_1} \frac{1}{2n} < L.$$
We continue the sequence (a_ℓ) by adjoining
$$(a_{m_1+1}, a_{m_1+2}, \ldots, a_{m_1+k_1}) = \left(-\frac{1}{2}, -\frac{1}{4}, \ldots, -\frac{1}{2k_1}\right).$$
Setting $m_2 = m_1 + k_1$, we have $s_{m_2} = P_1 + N_1 < L$.

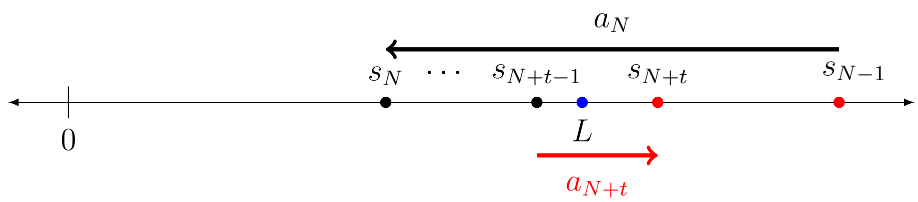
(3) Now add up just enough of the next unused terms from series P to obtain a sum P_2 satisfying $s_{m_2} + P_2 > L$. Adjoining these positive terms to the portion of the sequence (a_ℓ) already constructed, we get $s_{m_3} = s_{m_2} + P_2 > L$.

(4) Keep going, hopping back and forth over L by alternating between the series P and N.

Note the following points about this construction:

- At each step we adjoin at least one unused term from either P or N. It follows that we eventually use every term of P and N exactly once, and so (a_ℓ) is a rearrangement of the alternating harmonic series.
- The sequence (a_ℓ) has a first positive string of terms, then a first negative string of terms, then a second positive string of terms, then a second negative string of terms, etc. At the end of the kth positive string of terms, we have certainly made use of every odd reciprocal up to and including $1/(2k-1)$, and so all later positive terms of (a_ℓ) are smaller than $1/2k$. Similarly, after the kth negative string of terms, we have made use of every even reciprocal up to and including $1/2k$, so all later negative terms of (a_ℓ) also have size smaller than $1/2k$. Hence, after the kth pair of positive and negative strings, all terms of (a_ℓ) are smaller than $1/2k$.
- Corresponding to the strings of positive and negative terms in (a_ℓ), the sequence of partial sums (s_m) has stretches of increase and decrease. Moreover, the sequence (s_m) changes direction just after it hops over L. Moreover, if we consider one of these hops $s_{m-1} < L < s_m = s_{m-1} + a_m$, then $|L - s_m| < |a_m|$. That is: when the partial sums hop over L, they overshoot by at most the size of the last used term, and then they head in the other direction, back toward L.

We now wish to show that the sequence of partial sums (s_m) converges to L. For this, we play the convergence game. You begin by choosing a positive distance $d > 0$, and I must respond with an index N such that $|s_m - L| < d$ whenever $m \geq N$. To find N, I first select an integer k such that $1/k < d$. Then consider the kth pair of positive and negative strings of (a_ℓ) just discussed. Let N be the index of the last term a_N of the kth negative string. As mentioned above, $|a_\ell| \leq 1/2k$ for every $\ell \geq N$. Now we must verify that my choice of N is valid. We will make use of the following picture:



To get started, consider the partial sum s_N, which is a turning point for the sequence:

$$s_{N-1} + a_N = s_N < L < s_{N-1}.$$

It follows that $|L - s_N| < |a_N| \leq 1/2k < d$. Beginning with s_N, the partial sums increase for a while until reaching the next turning point s_{N+t}. All the partial sums $s_{N+1}, s_{N+2}, \ldots, s_{N+t-1}$ are between s_N and L, hence closer to L than d. For the turning point s_{N+t} we have

$$s_{N+t-1} < L < s_{N+t} = s_{N+t-1} + a_{N+t} < s_{N+t-1} + \frac{1}{2k} < s_{N+t-1} + d.$$

Thus, we have $|s_{N+t} - L| < d$ as well. Continuing in this way, passing from turning point to turning point, we see that all partial sums s_m for $m \geq N$ are within a distance d of L. Thus, my choice of N is valid, and we have proved that $\lim_{m\to\infty} s_m = L$, which means that

$$\sum_{\ell=1}^{\infty} a_\ell = L.$$

□

It turns out that this strange behavior is not confined to the alternating harmonic series, but occurs for all real conditionally convergent series. Happily, it cannot occur for absolutely convergence series. These facts are recorded in the following theorem, due to Riemann:

Theorem 3.62 (Riemann Rearrangement). *Consider a convergent real series $\sum_{n=1}^{\infty} b_n = s$.*

(1) If $\sum_{n=1}^{\infty} b_n$ is absolutely convergent, then every rearrangement (a_ℓ) of the sequence (b_n) sums to the same value s:

$$\sum_{\ell=1}^{\infty} a_\ell = \sum_{n=1}^{\infty} b_n = s.$$

(2) *If* $\sum_{n=1}^{\infty} b_n$ *is conditionally convergent, then for every real number* L*, there exists a rearrangment* (a_ℓ) *of the sequence* (b_n) *that sums to* L*:*

$$\sum_{\ell=1}^{\infty} a_\ell = L.$$

We finish this section with a complex version of this result. It was discovered by the 19 year-old Paul Lévy in 1905 while he was still an undergraduate at the École Polytechnique in Paris. He actually formulated his result for series of vectors in $\mathbb{R}^d$, but there was a gap in his proof for $d > 2$ that was filled in 1913 by the German mathematician Ernst Steinitz. The case $d = 2$ applies to the complex numbers $\mathbb{C}$:

THEOREM 3.63 (Lévy Rearrangement). *Suppose that* $\sum_{n=1}^{\infty} c_n = s$ *is a convergent complex series.*

(1) *If* $\sum_{n=1}^{\infty} c_n$ *is absolutely convergent, then every rearrangement* (a_ℓ) *of the sequence* (c_n) *sums to the same value* s*:*

$$\sum_{\ell=1}^{\infty} a_\ell = \sum_{n=1}^{\infty} c_n = s.$$

(2) *If* $\sum_{n=1}^{\infty} c_n$ *is conditionally convergent, then let* S *denote the collection of all sums of rearrangements of* (b_n)*. That is: a complex number* z *is in the set* S *if and only if there exists a rearrangement* (a_ℓ) *of the sequence* (b_n) *such that* $\sum_{\ell=1}^{\infty} a_\ell = z$*. Then there are two possibilities:*
(a) *the set* S *is a line in the complex plane;*
(b) *the set* $S = \mathbb{C}$ *is the entire complex plane.*

Key points for Section 3.8:

- Riemann Rearrangement Theorem (Theorem 3.62)

3.9. In-text Exercises

This section collects the in-text exercises that you should have worked on while reading the chapter.

EXERCISE 3.1 What is wrong with the following telescoping argument that purports to show that the series sums to 0?

$$\begin{aligned}\sum_{n=1}^{\infty}(-1)^{n+1} &= 1-1+1-1+1-1+\cdots \\ &= (1-1)+(1-1)+(1-1)+\cdots \\ &= 0+0+0+\cdots \\ &= 0.\end{aligned}$$

Can you make a similar (invalid) argument that suggests that the series sums to 1?

EXERCISE 3.2 Prove parts (*b*) and (*c*) of Proposition 3.11.

EXERCISE 3.3 Use Proposition 2.25 to prove Proposition 3.12.

EXERCISE 3.4 After reading the proof of the integral test, write a paragraph explaining where each of the hypotheses on the function f are used in the proof: (1) continuous, (2) nonnegative, (3) decreasing.

EXERCISE 3.5 Fill in the details of the proof of part (2) of the ratio test, Proposition 3.39.

EXERCISE 3.6 Finish the proof of the limit comparison test (Proposition 3.46) by showing that the convergence of $\sum_{n=1}^{\infty} a_n$ implies the convergence of $\sum_{n=1}^{\infty} b_n$.

EXERCISE 3.7 Give an example of a sequence (a_n) that has factorial growth.

EXERCISE 3.8 Suppose that $d > c > 1$, and show that d-exponential growth dominates c-exponential growth:

$$\lim_{n\to\infty}\frac{c^n}{d^n} = 0.$$

EXERCISE 3.9 Finish the proof of Proposition 3.58 by modifying the end of the proof of Proposition 3.57.

3.10. Problems

3.1. For the following series, compute the m-th partial sum for $m = 1, 2, 3, 4$:

(a) $\sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^k$
(b) $\sum_{k=1}^{\infty} \left(\frac{-2}{3}\right)^k$
(c) $\sum_{k=1}^{\infty} \left(\frac{2i}{3}\right)^k$
(d) $\sum_{n=1}^{\infty} \frac{1}{n(n+2)}$
(e) $\sum_{n=1}^{\infty} \frac{(-1)^n}{n(n+2)}$
(f) $\sum_{n=1}^{\infty} \frac{n}{n+1}$
(g) $\sum_{n=1}^{\infty} (2i)^n$

3.2. Consider the infinite series $\sum_{n=1}^{\infty} \frac{1}{n(n+2)}$.

(a) Find real numbers A and B with $\frac{1}{n(n+2)} = \frac{A}{n} + \frac{B}{n+2}$.
(b) Use a telescoping argument to show that $\sum_{n=1}^{\infty} \frac{1}{n(n+2)}$ converges and find the sum of the series.

3.3. Consider the infinite series $\sum_{n=1}^{\infty} \frac{1}{n(n+3)}$.

(a) Find real numbers A and B with $\frac{1}{n(n+3)} = \frac{A}{n} + \frac{B}{n+3}$.
(b) Use a telescoping argument to show that $\sum_{n=1}^{\infty} \frac{1}{n(n+3)}$ converges and find the sum of the series.

3.4. Consider the infinite series $\sum_{n=1}^{\infty} \frac{1}{4n^2-1}$.

(a) Find real numbers A and B with $\frac{1}{4n^2-1} = \frac{A}{2n-1} + \frac{B}{2n+1}$.
(b) Use a telescoping argument to show that $\sum_{n=1}^{\infty} \frac{1}{4n^2-1}$ converges and find the sum of the series.

3.5. Let $a_n = \frac{3n}{4n-1}$.

(a) Determine whether the sequence (a_n) converges or diverges.
(b) Determine whether the series $\sum_{n=0}^{\infty} a_n$ converges or diverges.

3.6. Let $b_n = (3i)^n$.

(a) Determine whether the sequence (b_n) converges or diverges.
(b) Determine whether the series $\sum_{n=0}^{\infty} b_n$ converges or diverges.

3.7. Let $c_n = \left(\frac{i}{3}\right)^n$.

(a) Determine whether the sequence (c_n) converges or diverges.
(b) Determine whether the series $\sum_{n=0}^{\infty} c_n$ converges or diverges.

3.8. Show that the following real series diverge:

(a) $\sum_{n=1}^{\infty} \frac{2n}{3n-1}$
(b) $\sum_{k=1}^{\infty} \left(\frac{4}{3}\right)^k$
(c) $\sum_{n=1}^{\infty} \cos\left(\frac{1}{n+1}\right)$
(d) $\sum_{n=1}^{\infty} \frac{n}{\sqrt{n^2+n-1}}$
(e) $\sum_{n=0}^{\infty} \left(\sqrt{4n^2+1}-n\right)$
(f) $\sum_{n=1}^{\infty} \frac{1}{\arctan(n)}$

3.9. Suppose that $\sum_{n=1}^{\infty} c_n$ is a convergent series with complex terms. Show that $\sum_{n=1}^{\infty} \overline{c_n}$ must also converge.

3.10. In Section 3.2, we learned that $\sum_{n=0}^{\infty} c^n = \frac{1}{1-c}$, when $|c| < 1$. Fix some c with $|c| < 1$.

(a) Show that
$$\sum_{n=1}^{\infty} c^n = \frac{c}{1-c}.$$
(b) Find the sum of the infinite series $\sum_{n=2}^{\infty} c^n$.

3.11. Determine whether the following geometric series converge or diverge. If the series converges, find its sum.

(a) $\sum_{n=0}^{\infty} \left(\frac{2i}{3}\right)^n$
(b) $\sum_{n=1}^{\infty} \left(\frac{3i}{2}\right)^n$
(c) $\sum_{n=1}^{\infty} \left(\frac{\pi}{3}\right)^n$
(d) $\sum_{n=1}^{\infty} \left(\frac{-5}{6}\right)^n$
(e) $\sum_{n=1}^{\infty} \left(\frac{-5i}{6}\right)^n$
(f) $\sum_{n=2}^{\infty} \left(\frac{3}{4}\right)^n$

3.12. Consider the infinite series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)}$.

(a) Find constants A, B, C such that $\frac{1}{n(n+1)(n+2)} = \frac{A}{n} + \frac{B}{n+1} + \frac{C}{n+2}$.
(b) Use a telescoping argument to show that $\sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)}$ converges and find the sum of the series.

3.13. For the following series, find all values of c for which the series converges. Draw a picture of the region on the complex plane.

(a) $\sum_{n=1}^{\infty} \frac{c^n}{3^n}$
(b) $\sum_{n=1}^{\infty} (c-2)^n$
(c) $\sum_{n=1}^{\infty} (2ic)^n$
(d) $\sum_{n=1}^{\infty} \left(\frac{3i}{4c}\right)^n$

3.14. Find all values of c for which
$$\sum_{n=2}^{\infty} (1+c)^{-n} = 1.$$

3.15. For each of the following series: (1) find a positive, decreasing, and continuous function $f(x)$ with $f(n) = a_n$ for all n, and (2) use the integral test to determine whether the series converges or diverges.

(a) $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n+3}}$
(b) $\sum_{n=2}^{\infty} \frac{1}{n\ln(n)}$
(c) $\sum_{n=1}^{\infty} \frac{n^2}{n^3+1}$

3.16. Use the integral test to find all values of p for which $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^p}$ converges.

3.17. Use the integral test to find all values of p for which $\sum_{n=1}^{\infty} n(1+n^2)^p$ converges.

3.18. Estimate $\sum_{n=1}^{\infty} \frac{1}{n^5}$ correct to three decimal places.

3.19. Estimate $\sum_{n=1}^{\infty}(2n+1)^{-6}$ correct to five decimal places.

3.20. Use the direct comparison test to determine whether the following real series converge or diverge:

(a) $\sum_{n=1}^{\infty} \frac{n-1}{n^2\sqrt{n}}$
(b) $\sum_{n=2}^{\infty} \frac{\sqrt{n}}{n-1}$
(c) $\sum_{n=1}^{\infty} \frac{3^n}{n+4^n}$
(d) $\sum_{n=1}^{\infty} \frac{\sin(n)+1}{n^2}$
(e) $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^3+n}}$
(f) $\sum_{n=1}^{\infty} \frac{\cos^2(n)}{2^n}$
(g) $\sum_{n=4}^{\infty} \frac{\sqrt{n}}{n-3}$

3.21. Determine whether the following alternating real series converge or diverge.

(a) $\sum_{n=2}^{\infty}(-1)^{n+1}\frac{1}{n\ln n}$
(b) $\sum_{n=1}^{\infty}(-1)^{n}\frac{\ln n}{n}$
(c) $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{n}{4^{n+1}}$
(d) $\sum_{n=1}^{\infty}(-1)^{n}\cos\left(\frac{1}{n}\right)$
(e) $\sum_{n=1}^{\infty}(-1)^{n}\sin\left(\frac{1}{n}\right)$

3.22. Determine whether the following series converges: $\sum_{n=1}^{\infty} \frac{i^n}{n}$.

3.23. Show that the following complex series are absolutely convergent:

(a) $\sum_{n=1}^{\infty} \frac{\sqrt{n}+in}{n^3}$

(b) $\sum_{n=1}^{\infty} \frac{(1+i)^n}{n^2 2^n}$

(c) $\sum_{n=1}^{\infty} \frac{\sqrt{n}-1}{\sqrt{n}+n^2 i}$

(d) $\sum_{n=1}^{\infty} \frac{(3+4i)^n}{5^n(n+n^3 i)}$

(e) $\sum_{n=1}^{\infty} \frac{(3+2i)^n}{n(4+i)^n}$

3.24. Use the ratio test to determine whether the following real series converge or diverge:

(a) $\sum_{n=1}^{\infty} \frac{2^n}{n^5}$

(b) $\sum_{n=0}^{\infty} \frac{\sqrt{n}}{3^n}$

(c) $\sum_{n=0}^{\infty} \frac{n^2}{(2n+1)!}$

(d) $\sum_{n=1}^{\infty} \frac{e^n}{(2n)!}$

(e) $\sum_{n=1}^{\infty} \frac{(n!)^2}{(2n)!}$

3.25. Use the ratio test to determine whether the following complex series converge or diverge.

(a) $\sum_{n=1}^{\infty} \frac{(2+i)^n}{n}$

(b) $\sum_{n=0}^{\infty} \frac{(3i)^n n}{4^n}$

(c) $\sum_{n=1}^{\infty} \frac{(2+3i)^n}{\pi^n n}$

(d) $\sum_{n=1}^{\infty} \frac{(1+i)^n \sqrt{n}}{3^n}$

(e) $\sum_{n=1}^{\infty} \frac{\sqrt{n+1}}{(5-i)^n}$

3.26. In Example 3.43, we showed that the series $\sum_{n=0}^{\infty} \frac{1}{n!}$ converges by using the ratio test. Provide an alternate proof of convergence for this series using the comparison test. (Hint: use the geometric series $\sum_{n=0}^{\infty} \frac{1}{2^n}$.)

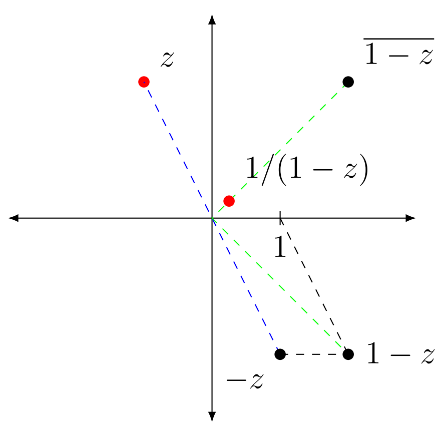
CHAPTER 4

POWER SERIES

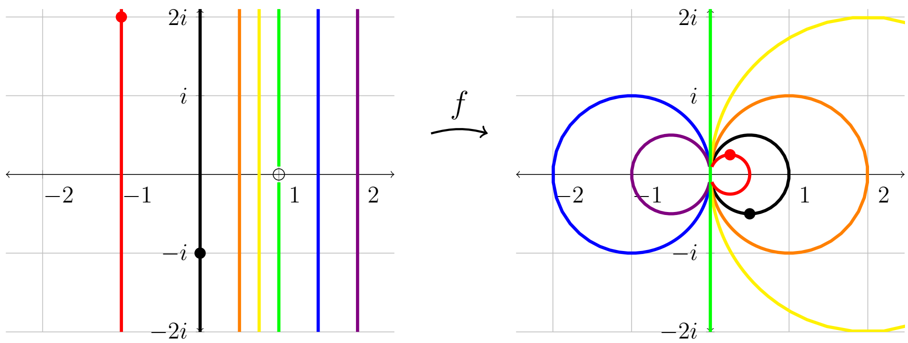
4.1. Preview: The Geometric Series

We preview the topics of this chapter by focusing on a single important example: the complex function defined by $f(z) = 1/(1-z)$. The domain of $f(z)$ is the entire complex plane except for the point $z = 1$, because this point would lead to a zero in the denominator of the expression $1/(1-z)$. Let's begin by tracking the effect of f on a single complex number z using the geometric ideas from Section 1.1:

$$z \mapsto -z \mapsto 1-z \mapsto \overline{1-z} \mapsto \frac{1}{1-z} = \frac{\overline{1-z}}{|1-z|^2}$$



With this in mind, consider the following picture showing the effect of the function f on vertical lines in the style of Section 1.3:



As in the previous picture, the pair of red dots represents the transformation of the point $z = -1 + 2i$ into $f(-1 + 2i) = \frac{1}{4} + \frac{i}{4}$:

$$\frac{1}{1-(-1+2i)} = \frac{1}{2-2i} = \frac{2+2i}{(2-2i)(2+2i)} = \frac{2+2i}{8} = \frac{1}{4} + \frac{i}{4}.$$

EXERCISE 4.1. Spend some time pondering the picture above.

(a) Check that $f(-i) = \frac{1}{2} - \frac{i}{2}$ as indicated by the pair of black dots in the picture.

(b) Now consider a general point $z = iy$ on the imaginary axis, so

$$f(iy) = \frac{1}{1-iy} = \frac{1+iy}{1+y^2}.$$

As y varies, do you see why $f(iy)$ traces out the black circle on the right hand side?

(c) Can you convince yourself that f takes vertical lines to circles, as the picture indicates?

We now want to make contact with our study of series. To begin, recall the convergence result established in Example 3.10: if c is a complex number with $|c| < 1$, then the geometric series with common ratio c converges:

$$\sum_{n=0}^{\infty} c^n = \frac{1}{1-c}.$$

This formula is valid for *all* complex numbers c in the open disc of radius 1 centered at the origin. Notice that the expression on the right hand side is $f(c)$, the value of our function at the point $z = c$. So the convergence result for geometric series provides a *series formula* for the function f on the restricted domain $|z| < 1$:

$$f(z) = \frac{1}{1-z} = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots$$

The change of notation from c to z emphasizes the shift in our focus compared with Chapter 3: instead of focusing on individual series of complex numbers $\sum_{n=0}^{\infty} c^n$, we now want to focus on the series formula $\sum_{n=0}^{\infty} z^n$ for the complex function $f(z)$.

But what is the meaning of this series formula for $f(z)$? Recall that the value of a convergent infinite series is given by the limit of its partial sums, and in this case the partial sums are polynomials in the complex variable z:

$$\begin{aligned} s_0(z) &= 1 \\ s_1(z) &= 1 + z \\ s_2(z) &= 1 + z + z^2 \\ s_3(z) &= 1 + z + z^2 + z^3 \\ &\vdots \end{aligned}$$

So, the series formula $f(z) = \sum_{n=0}^{\infty} z^n$ expresses the fact that, on the domain $|z| < 1$, the function $f(z)$ can be expressed as the limit of polynomials:

$$f(z) = \frac{1}{1-z} = \lim_{m\to\infty} s_m(z) = \lim_{m\to\infty} (1 + z + z^2 + z^3 + \cdots + z^m).$$

We would like to get a visual sense of what it means for the function $f(z)$ to be the limit of polynomials, and for this it will be extremely helpful to restrict attention to real values $z = x$, so that we can look at graphs of real functions. So, we now consider the real function $f(x) = 1/(1-x)$ and the real partial sums $s_m(x) = 1+x+x^2+\cdots+x^m$. On the domain $|x| < 1$, we have

$$f(x) = \frac{1}{1-x} = \lim_{m\to\infty} s_m(x) = \lim_{m\to\infty} (1 + x + x^2 + x^3 + \cdots + x^m).$$

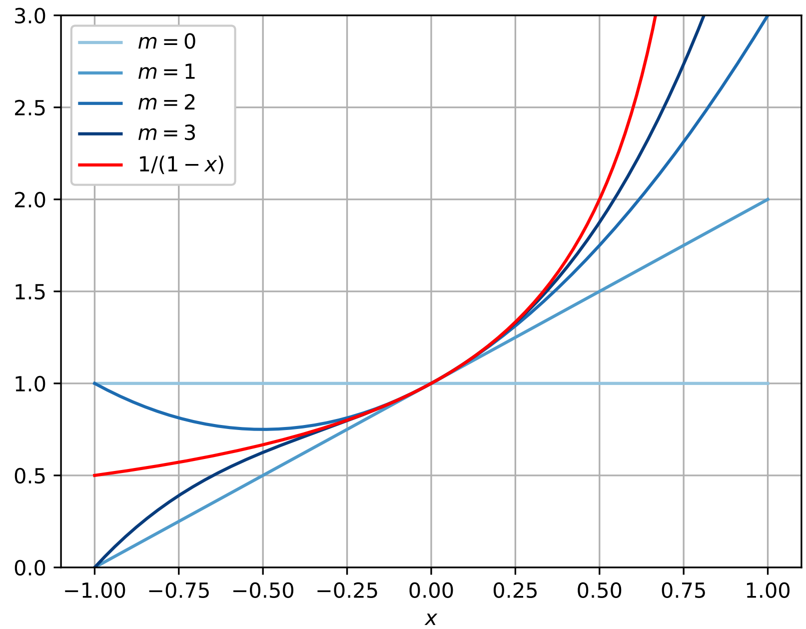


FIGURE 4.1. The first four polynomial partial sums $s_m(x)$ for the geometric series representation of $1/(1-x)$.

Figure 4.1 shows the graph of $f(x)$ together with graphs of the first few polynomial partial sums. We interpret the figure as follows:

- the partial sum $s_0(x) = 1$ (lightest blue) is the best constant approximation of $f(x) = 1/(1-x)$ (red) near the point $x = 0$;
- the partial sum $s_1(x) = 1 + x$ (light blue) is the best linear approximation: its graph is the tangent line to $f(x)$ at $x = 0$;
- the next partial sum $s_2(x)$ (medium blue) appears to be the best quadratic approximation to $f(x)$ at $x = 0$;
- the degree-3 polynomial $s_3(x)$ (darkest blue) seems to be the best cubic approximation.

To explain this terminology, recall that the tangent line to the graph of a function f at a point c provides the best linear approximation to the graph at c, because it shares both the value of the function $f(c)$ and the slope $f'(c)$ at that point. Generalizing this idea, the best quadratic approximation at c will share not only the value $f(c)$ and the slope $f'(c)$, but it will also have the same second derivative $f''(c)$. In general, when we say that the polynomial $s_m(x)$ is the *best mth degree approximation* of $f(x)$ near $x = c$, we mean that $s_m(x)$ not only has the same value as f at $x = c$, but also the same first, second, ..., mth derivatives:

$$s_m(c) = f(c), \quad s_m'(c) = f'(c), \quad s_m''(c) = f''(c), \ldots, s_m^{(m)}(c) = f^{(m)}(c).$$

EXERCISE 4.2. For $f(x) = 1/(1 - x)$ and the third partial sum $s_3(x) = 1 + x + x^2 + x^3$, verify by explicit computation that

$$f(0) = s_3(0), \quad f'(0) = s_3'(0), \quad f''(0) = s_3''(0), \quad f^{(3)}(0) = s_3^{(3)}(0).$$

Figure 4.1 helps us understand the meaning of the series formula

$$f(x) = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n.$$

Namely, the partial sums $s_m(x)$ provide better and better approximations to the function $f(x)$ near the point $x = 0$, and in the limit as $m \to \infty$, these polynomial approximations converge to provide an exact match for $f(x)$ on the domain $|x| < 1$. Moreover, even though we cannot easily visualize the graphs of the complex function $f(z) = 1/(1-z)$ and the complex partial sums $s_m(z)$, the meaning of the complex series formula $f(z) = \sum_{n=0}^{\infty} z^n$ is the same: the complex polynomials $s_m(z)$ provide approximations to the complex function $f(z)$, and in the limit as $m \to \infty$, the polynomials converge to provide an exact match for $f(z)$ on the open disc $|z| < 1$.

REMARK 4.1. Note that the domain of the function $f(z) = 1/(1 - z)$ is larger than the domain of convergence $|z| < 1$ for the geometric series formula $\sum_{n=0}^{\infty} z^n$. Thus, while the geometric series does represent the function $f(z)$, it does so only on a small portion of the original domain of f.

Based on this example, we now provide a list of questions that will serve as a roadmap for the chapter ahead:

(1) What are the domains of convergence for series formulas of the type $F(z) = \sum_{n=0}^{\infty} a_n z^n$, where z is a complex variable and the coefficients a_n are complex numbers? Such functions are called *power series*, and they are the main topic of this chapter.
(2) We know from calculus that real polynomials are nice functions: they are differentiable, and their derivatives are easy to compute term-by-term using the power rule. What about complex polynomials?
(3) Power series are limits of polynomials, so we wonder: do the nice properties of polynomials carry over to power series?
(4) Given a complex function $G(z)$, is there a power series formula $G(z) = \sum_{n=0}^{\infty} a_n z^n$, valid on some domain in the complex plane? If so, how can we find the numerical coefficients a_n explicitly?
(5) Given a real function $g(x)$, can power series help us to find a nice complex function $G(z)$ that extends g to the complex plane? That is, can we find a complex function $G(z)$ such that $G(x) = g(x)$ for real numbers x? We will be especially interested in extending the familiar trigonmetric functions $\cos(x)$ and $\sin(x)$ to the complex plane, as well as the exponential function e^x.
(6) What are some applications of power series?

Key points from Section 4.1:

- The geometric series formula for $f(z) = 1/(1-z)$:

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n, \qquad |z| < 1.$$

- $f(z)$ as a limit of polynomials (Figure 4.1)
- Best mth degree polynomial approximation (page 177)

4.2. The Radius of Convergence

We begin by giving a precise definition of the power series mentioned in the previous section.

DEFINITION 4.2. A *power series* is a series of the form

$$\sum_{n=0}^{\infty} a_n z^n = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots .$$

Here, z is a complex variable, and the coefficients a_n are complex numbers. This series will generally converge for some values of z and diverge for other values. Letting D denote the set of complex numbers z for which the series converges, we obtain a complex function $F\colon D \to \mathbb{C}$ defined by the power series:

$$F(z) = \sum_{n=0}^{\infty} a_n z^n.$$

EXAMPLE 4.3. Generalizing the example from the previous section, for every complex number a, the geometric series with first term a defines a power series on the domain $|z| < 1$:

$$F(z) = \sum_{n=0}^{\infty} az^n = a + az + az^2 + az^3 + \cdots = \frac{a}{1-z}.$$

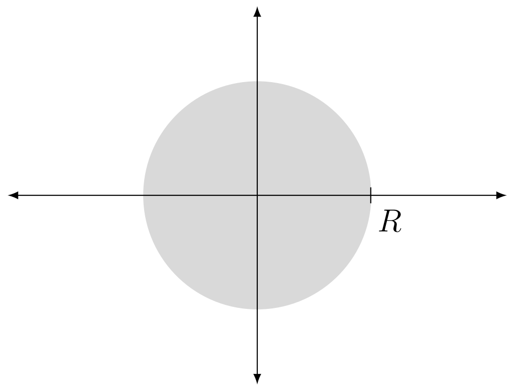
A caution about domains: the formula $a/(1-z)$ on the right hand side makes sense for all $z \neq 1$, but the series converges only on the much smaller domain $|z| < 1$. As a concrete example, consider the point $z = 3$. Then the right hand side makes perfect sense: $a/(1-3) = -a/2$. But the left hand side is a divergent geometric series:

$$\sum_{n=0}^{\infty} a \cdot 3^n.$$

So: the function $F(z)$ defined by the power series has domain $|z| < 1$, even though in this special case we see how to extend the function to a larger domain by using the formula $a/(1-z)$ on the right hand side.

Note that geometric series define functions with especially nice domains: the power series $\sum_{n=0}^{\infty} az^n$ converges on the open unit disc

$|z| < 1$. In fact, the next proposition shows that all power series have nice domains.



PROPOSITION 4.4. *Suppose that* $\sum_{n=0}^{\infty} a_n z^n$ *is a power series. Then there is a real number* $R \geq 0$ *(or* $R = +\infty$*) such that the power series is absolutely convergent for* $|z| < R$ *and divergent for* $|z| > R$.

DEFINITION 4.5. The number R in Proposition 4.4 is called the *radius of convergence* of the power series. The picture above indicates the reason for this term: the power series converges inside the disc of radius R.

REMARK 4.6. We will prove Proposition 4.4 under the additional assumptions that (1) the coefficients a_n are eventually nonzero and (2) the magnitudes of the ratios of successive coefficients converge to a real number L or diverge to $+\infty$:

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L \qquad \text{or} \qquad \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = +\infty.$$

These extra hypotheses are not actually necessary, but assuming them allows us to give an easier proof that explicitly identifies the radius of convergence as $R = 1/L$ and avoids a more technical argument involving completeness. Moreover, these extra hypotheses will hold for all the examples in this course (although see Example 4.9 below for more about the assumption (1) that the coefficients a_n are eventually nonzero).

PROOF. First note that the series certainly converges (to a_0) for $z = 0$. So assume that $|z| > 0$ and use the ratio test:

$$\lim_{n\to\infty} \frac{|a_{n+1}z^{n+1}|}{|a_n z^n|} = \lim_{n\to\infty} |z|\frac{|a_{n+1}|}{|a_n|} = \begin{cases} |z|L \\ +\infty, \end{cases}$$

according to which case of our extra assumption (2) holds. If the ratios diverge to $+\infty$, then the ratio test says the power series diverges for all nonzero z, and we take $R = 0$. On the other hand, if $L = 0$, then the ratio test says that the power series converges absolutely for all z, and we take $R = +\infty$. Finally, if $0 < L < +\infty$, then we set $R = 1/L$. In this case the ratio test says that the series converges absolutely if $|z| < R = 1/L$ (since then $|z|L < 1$) and diverges if $|z| > R = 1/L$ (since then $|z|L > 1$). □

EXAMPLE 4.7. Consider the power series with coefficients $a_n = 2^n\sqrt{n}$:

$$\sum_{n=0}^{\infty} 2^n\sqrt{n}z^n = 2z + 4\sqrt{2}z^2 + 8\sqrt{3}z^3 + \cdots$$

To find the radius of convergence, we use the ratio test as in the proof of Proposition 4.4:

$$\begin{aligned}
\lim_{n\to\infty} \frac{|2^{n+1}\sqrt{n+1}z^{n+1}|}{|2^n\sqrt{n}z^n|} &= \lim_{n\to\infty} \frac{2^{n+1}}{2^n} \cdot \sqrt{\frac{n+1}{n}} \cdot \frac{|z|^{n+1}}{|z|^n} \\
&= \lim_{n\to\infty} 2 \cdot \sqrt{1 + \frac{1}{n}} \cdot |z| \\
&= 2|z| \lim_{n\to\infty} \sqrt{1 + \frac{1}{n}} \\
&= 2|z|.
\end{aligned}$$

For the ratio test to guarantee absolute convergence, we must have $2|z| < 1$ or $|z| < 1/2$, which tells us that the radius of convergence is $R = 1/2$.

EXAMPLE 4.8. Now consider the power series with factorial coefficients $a_n = n!$:

$$\sum_{n=0}^{\infty} n!z^n = 1 + z + 2z^2 + 6z^3 + 24z^4 + 120z^5 + \cdots$$

We use the ratio test to find the radius of convergence:

$$\begin{aligned}\lim_{n\to\infty}\frac{|(n+1)!z^{n+1}|}{|n!z^n|} &= \lim_{n\to\infty}\frac{(n+1)!}{n!}\cdot\frac{|z|^{n+1}}{|z|^n}\\ &= \lim_{n\to\infty}(n+1)\cdot|z|\\ &= |z|\lim_{n\to\infty}(n+1)\\ &= +\infty.\end{aligned}$$

Thus, the ratios diverge to $+\infty$ for all nonzero z, so the power series converges only for $z=0$ and the radius of convergence is $R=0$. This example provides an illustration of the extremely fast growth rate of the factorial sequence $(n!)$. For more on that topic, see the optional Section 3.7.

Example 4.9. Consider the power series

$$\begin{aligned}\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n}}{(2n)!} &= 1-\frac{z^2}{2}+\frac{z^4}{4!}-\frac{z^6}{6!}+\cdots\\ &= 1+0\cdot z-\frac{z^2}{2}+0\cdot z^3+\frac{z^4}{4!}+0\cdot z^5-\frac{z^6}{6!}+\cdots\end{aligned}$$

The coefficients are $a_0=1$, $a_1=0$, $a_2=-1/2$, $a_3=0,\ldots$. In particular, all the terms with odd index are zero. At first glance, it would seem that we cannot use the ratio test to determine the radius of convergence, since the ratio $|a_{n+1}|/|a_n|$ is not defined for odd values of n. But note that if we set $w=z^2$, then we can rewrite the power series in terms of w, with nonzero coefficients $b_n=(-1)^n/(2n)!$:

$$\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n}}{(2n)!}=\sum_{n=0}^{\infty}(-1)^n\frac{w^n}{(2n)!}.$$

Now we find that

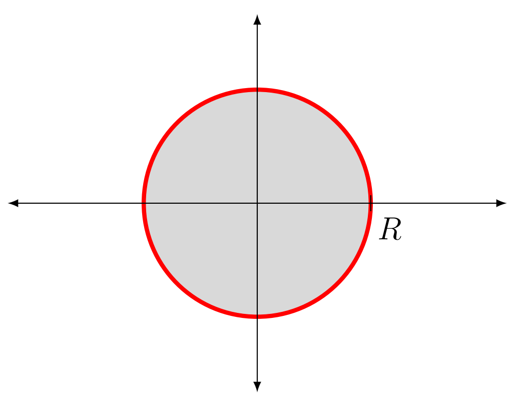
$$\lim_{n\to\infty}\frac{|b_{n+1}|}{|b_n|}=\lim_{n\to\infty}\frac{(2n)!}{(2n+2)!}=\lim_{n\to\infty}\frac{1}{(2n+2)(2n+1)}=0.$$

As in the proof of Proposition 4.4, the limit $L=0$ for the ratios of the coefficients implies that $R=+\infty$ and the series converges for all w. But since $w=z^2$, it follows that the original power series converges for all z. In this way, the ratio test can sometimes be used for series with "missing terms" (but see the remark below for an example where it cannot be used).

EXERCISE 4.3. Adapt the argument given in Example 4.9 to show that the power series below has radius of convergence $R = +\infty$.

$$\sum_{n=0}^{\infty}(-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots .$$

REMARK 4.10. Note that Proposition 4.4 says nothing about the behavior of a power series for $|z| = R$, shown as the red boundary circle below. This *boundary behavior* is quite delicate, and all sorts of things can happen: the series may converge nowhere or everywhere on the boundary circle, or it might converge at some points and diverge at others. Moreover, the convergence may only be conditional on the boundary. In this course, we will mainly ignore the issue of boundary behavior, and focus on the interior of the disc of radius R, where the power series converges absolutely. The next example illustrates some interesting boundary behavior.



EXAMPLE 4.11. Consider the power series

$$F(z) = \sum_{n=0}^{\infty} \frac{z^{(2^n)}}{2^n} = z + \frac{z^2}{2} + \frac{z^4}{4} + \frac{z^8}{8} + \cdots$$

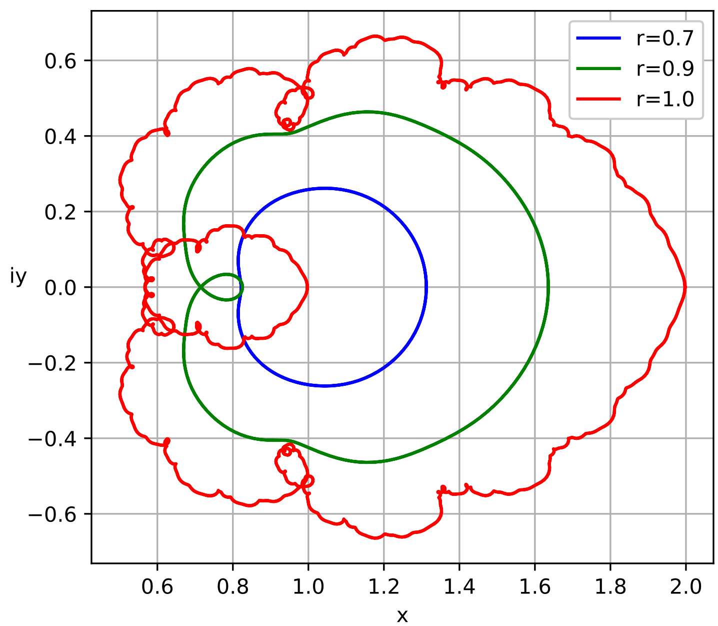


FIGURE 4.2. Plots showing the effect of the power series $F(z) = \sum_{n=0}^{\infty} \frac{z^{(2^n)}}{2^n}$ on three concentric circles of radius r centered at the origin. The red curve is the result of applying F to the circle $|z| = 1$, which forms the boundary of its domain.

This series has lots of "missing terms," and the gaps between nonzero terms grows larger and larger. For this reason, we can't use the method of Example 4.9 to determine the radius of convergence. But one can show (Problem 4.7) that the radius of convergence is $R = 1$, and moreover that the series converges absolutely on the disc $|z| \leq 1$. Figure 4.2 shows the effect of the power series $F(z)$ on some circles centered at the origin, including the boundary circle $|z| = 1$.

Key points from Section 4.2:

- The radius of convergence R (Definition 4.5)
- Using the ratio test to find the radius of convergence (Examples 4.7–4.9)

4.3. The Complex Derivative

In your first calculus course, you defined the derivative of a real function $f\colon \mathbb{R} \to \mathbb{R}$ at a point c by using function limits: the real function f is *differentiable at* c when the following limit exists:

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

The number $f'(c)$ is called the *derivative of* f *at* c.

You went on to establish many properties of differentiation, such as the sum, product, quotient, and chain rules. In particular, you proved that all real polynomials are differentiable, and their derivatives are easy to compute term-by-term using the power rule:

$$\frac{d}{dx}(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1}.$$

Our aim in this section is to extend the idea of differentiation to the complex setting, and to show that complex polynomials are also differentiable using the power rule. In order to do this, we begin by recasting our definition of the real derivative in terms of sequence limits rather than function limits.

In your first calculus course, you likely discussed the limit defining the derivative $f'(c)$ in the following informal way: the difference quotient $\frac{f(c+h)-f(c)}{h}$ gets arbitrarily close to the number $f'(c)$ as the real number h gets sufficiently close to 0. When thinking about this limit, we can imagine h approaching zero from the left or from the right. But no matter how h approaches zero, the corresponding difference quotients must approach the same number $f'(c)$; this is what it means for the function limit to exist. Figure 4.3 shows the graphical interpretation of convergence from the left and from the right by displaying secant lines converging to the tangent line for f at c.

This way of thinking—imagining all the different ways h can approach zero—suggests the following rewriting of the definition of real differentiability in terms of sequences (h_n) converging to zero:

Definition 4.12 (Sequential Definition of Real Derivative). Consider a real function $f(x)$ defined on an open interval containing the

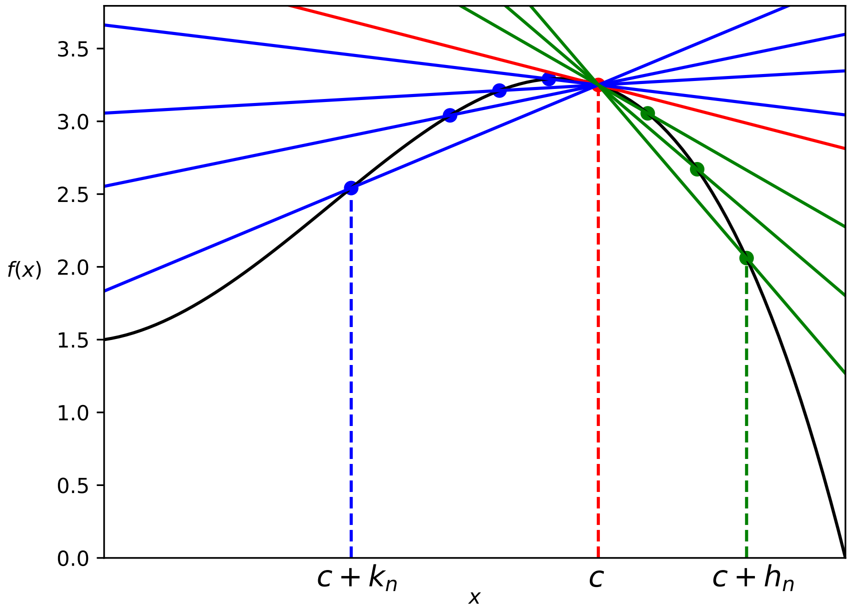


FIGURE 4.3. Plot of the graph of a function $f(x)$ together with its red tangent line at c and two families of secant lines. The green secants correspond to a positive sequence (h_n) converging to zero from the right, and the blue secants correspond to a negative sequence (k_n) converging to zero from the left. Differentiability of f at c means that the slopes of the blue lines converge to the same value as the slopes of the green lines, the common value being the slope of the red tangent line at c.

real number c. Then f is *differentiable at* c if for every sequence (h_n) of nonzero real numbers converging to zero, the following limit exists and is independent of the choice of sequence:

$$\lim_{n\to\infty} \frac{f(c+h_n) - f(c)}{h_n}.$$

In this case, we denote the common limit by $f'(c)$ and call it the *derivative of* f *at* c.

Now that we have a definition for the real derivative in terms of sequence limits, we can immediately promote it to the complex setting:

DEFINITION 4.13 (Sequential Definition of Complex Derivative). Suppose that $f(z)$ is a complex function defined on an open disc containing the complex number c. Then f is *differentiable* at c if for every sequence of nonzero complex numbers (h_n) converging to zero, the following limit exists and is independent of the choice of sequence:

$$\lim_{n\to\infty} \frac{f(c+h_n)-f(c)}{h_n}.$$

In this case, we denote the common limit by $f'(c)$ or $\frac{df}{dz}(c)$ and call it the *derivative of f at c*.

No matter what sequence (h_n) we consider, the limit of the corresponding difference quotients must exist and be equal to *the same* value $f'(c)$. Compared to the real setting, there are many more ways to approach zero in the complex plane (see Figure 4.4), and so complex differentiability is harder to achieve than real differentiability.

EXAMPLE 4.14. Consider the squaring function $f(z) = z^2$ and any point c in the complex plane. We want to investigate the existence of the derivative $f'(c)$. So let (h_n) be any sequence of nonzero complex numbers converging to zero, and look at the difference quotients:

$$\begin{aligned} \frac{f(c+h_n)-f(c)}{h_n} &= \frac{(c+h_n)^2-c^2}{h_n} \\ &= \frac{c^2+2ch_n+h_n^2-c^2}{h_n} \\ &= \frac{2ch_n+h_n^2}{h_n} \\ &= 2c+h_n. \end{aligned}$$

Having made this computation, we now take the limit:

$$\lim_{n\to\infty} \frac{f(c+h_n)-f(c)}{h_n} = \lim_{n\to\infty}(2c+h_n) = 2c + \lim_{n\to\infty} h_n = 2c.$$

Note the key point: the limit not only exists, but its value $2c$ does not depend on the sequence (h_n). This shows that $f(z) = z^2$ is differentiable at c with $f'(c) = 2c$. Since this is true for all complex numbers c,

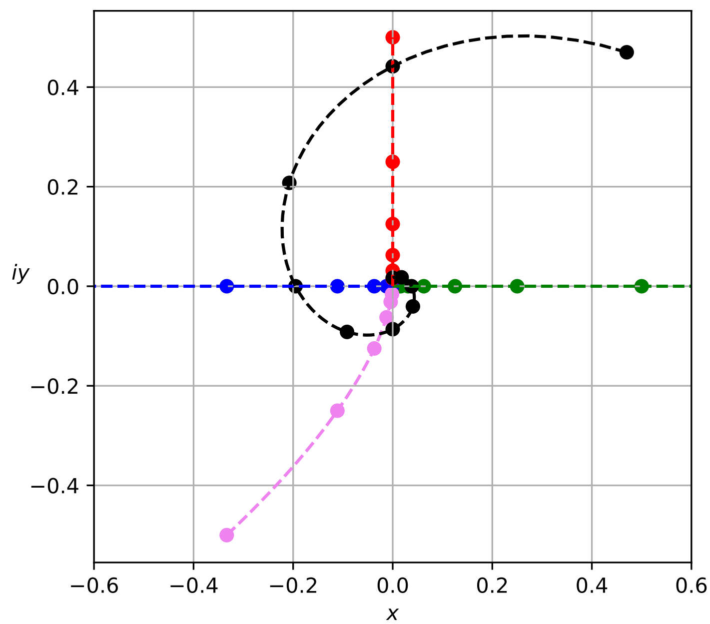


FIGURE 4.4. The blue sequence is approaching zero from the left along the real axis, while the green sequence is approaching from the right. The red sequence is approaching from above along the imaginary axis. The pink sequence is approaching along a slightly curved path in the 3rd quadrant, and the black sequence is spiraling inward.

we write $f'(z) = 2z$ or

$$\frac{d}{dz}(z^2) = 2z.$$

Here is an example showing the failure of complex differentiability for a seemingly nice function.

EXAMPLE 4.15. Consider the complex conjugation function $f(z) = \overline{z}$, defined by $f(x+iy) = x-iy$. Geometrically, this is the reflection across the real axis. Fix any complex number c, and consider the real sequence $(h_n) = (1/n)$, which approaches zero from the right. We find that

$$\frac{f(c+h_n) - f(c)}{h_n} = \frac{\overline{c+1/n} - \overline{c}}{1/n} = \frac{\overline{c} + 1/n - \overline{c}}{1/n} = 1.$$

On the other hand, if we use the imaginary sequence $(h_n) = (i/n)$ which approaches zero from above, we find

$$\frac{f(c+h_n)-f(c)}{h_n} = \frac{\overline{c+i/n}-\overline{c}}{i/n} = \frac{\overline{c}-i/n-\overline{c}}{i/n} = -1.$$

Thus, the limits for these two sequences do not agree, and so complex conjugation is not complex differentiable at any point!

Despite the lesson you might draw from the previous example, there are many complex differentiable functions. In particular, the same proof that you saw for the real function x^m in your first calculus course works to establish the power rule for the derivative of the complex function z^m.

PROPOSITION 4.16 (Power Rule). *For any integer $m \geq 1$, the complex function z^m is differentiable, with derivative*

$$\frac{d}{dz}(z^m) = mz^{m-1}.$$

PROOF. For any complex number h we have

$$(z+h)^m = z^m + mhz^{m-1} + \binom{m}{2}h^2z^{m-2} + \cdots + mh^{m-1}z + h^m.$$

It follows that, for any sequence of nonzero complex numbers (h_n) converging to zero, the difference quotients are

$$\begin{aligned}\frac{(z+h_n)^m - z^m}{h_n} &= \frac{mh_nz^{m-1} + \binom{m}{2}h_n^2z^{m-2} + \cdots + mh_n^{m-1}z + h_n^m}{h_n} \\ &= mz^{m-1} + \binom{m}{2}h_nz^{m-2} + \cdots + mh_n^{m-2}z + h_n^{m-1}.\end{aligned}$$

The key point to notice is that every term except the first in this expression contains h_n, and so converges to zero as $n \to \infty$:

$$\begin{aligned}\lim_{n\to\infty}\frac{(z+h_n)^m - z^m}{h_n} &= \lim_{n\to\infty}\left(mz^{m-1} + \binom{m}{2}h_nz^{m-2} + \cdots + h_n^{m-1}\right) \\ &= mz^{m-1}.\end{aligned}$$

Since the limit exists and equals mz^{m-1} independently of the sequence (h_n), we find that z^m is differentiable with

$$\frac{d}{dz}(z^m) = mz^{m-1}.$$

□

More generally, all the usual differentiation rules hold for the complex derivative (the sum rule, the product rule, the quotient rule, and the chain rule), and their proofs are basically the same as in the real case. In particular, the derivative of a complex polynomial may be computed by using the power rule term-by-term:

PROPOSITION 4.17. *Complex polynomials are differentiable, and their derivatives may be computed term-by-term using the power rule:*

$$\frac{d}{dz}(a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n) = a_1 + 2a_2 z + \cdots + na_n z^{n-1}.$$

PROOF. This follows by combining the sum rule, the constant multiple rule, and the power rule. □

EXERCISE 4.4. Fill in the details in the proof of Proposition 4.17.

To illustrate the use of the chain and quotient rules, we prove the following result that we will use later.

PROPOSITION 4.18. *Suppose that m is an integer exponent and b a fixed complex number. Then the function $f(z) = (z-b)^m$ is complex differentiable, with*

$$f'(z) = m(z-b)^{m-1}.$$

PROOF. First consider the case where $m \geq 0$. By the chain rule and the power rule,

$$f'(z) = m(z-b)^{m-1}(z-b)' = m(z-b)^{m-1},$$

since $(z-b)' = 1$ by the previous proposition.

Now suppose that $m < 0$ is negative, and set $k = -m > 0$. Then the derivative of $(z-b)^k$ is $k(z-b)^{k-1}$ by the nonnegative case. But

$f(z) = 1/(z-b)^k$, and we can use the quotient rule:

$$\begin{aligned} f'(z) &= \frac{0 \cdot (z-b)^k - 1 \cdot k(z-b)^{k-1}}{(z-b)^{2k}} \\ &= \frac{-k}{(z-b)^{k+1}} \\ &= m(z-b)^{m-1}. \end{aligned}$$

□

Key points from Section 4.3:

- Definition of complex derivative (Definition 4.13)
- Term-by-term differentiation of polynomials (Proposition 4.17)

4.4. Power Series as Infinite Polynomials

Now consider a power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$, with radius of convergence $R > 0$. As discussed in Section 4.1, the partial sums are polynomials, so the function $f(z)$ is a limit of polynomials:

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \lim_{n\to\infty} (a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n).$$

Since the function $f(z)$ is the limit of differentiable polynomials, it seems reasonable to expect that $f(z)$ should be differentiable. This is actually true for power series, although the proof is technical and we will not give it here—it is better suited for a later course in analysis. However, in other contexts besides power series, the result may not hold! That is, there are examples of non-differentiable functions $g(z) = \lim_{n\to\infty} g_n(z)$ that are limits of differentiable functions $g_n(z)$. Figure 4.5 indicates what can go wrong.

But in the special case of power series, the convergence of the polynomial partial sums is of a good type (called *uniform*), and we have the following extremely useful result:

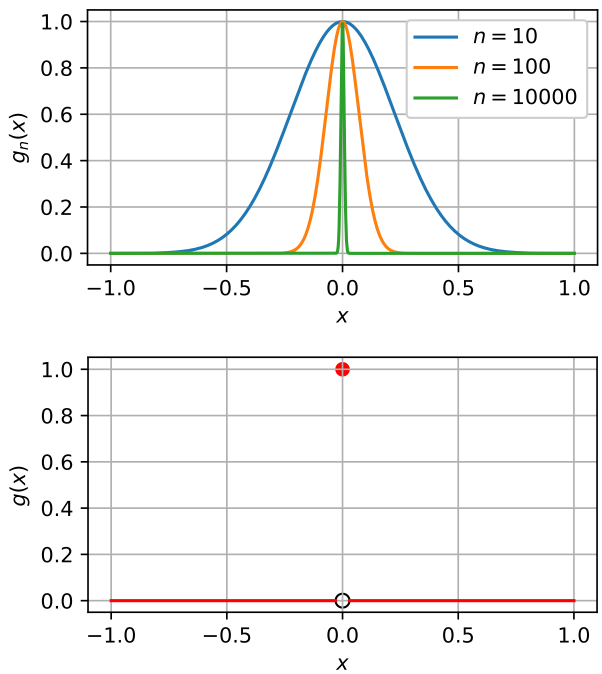


FIGURE 4.5. The upper plot shows the differentiable functions $g_n(x) = e^{-nx^2}$; each is a smooth bell shape of height 1, but they get narrower and narrower as n increases. The lower plot shows the limit function $g(x) = \lim_{n\to\infty} g_n(x)$; it is zero except at $x = 0$, where it has a jump discontinuity.

THEOREM 4.19. *Suppose that $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is a power series with radius of convergence $R > 0$. Then the function $f(z)$ is differentiable on the open disc of radius R, and its derivative is given by term-by-term differentiation using the power rule:*

$$f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1} = a_1 + 2a_2 z + 3a_3 z^2 + \cdots .$$

Moreover, the power series defining $f'(z)$ has the same radius of convergence R.

EXAMPLE 4.20. Consider the power series from Example 4.7, which has radius of convergence $R = 1/2$:

$$f(z) = \sum_{n=0}^{\infty} 2^n \sqrt{n} z^n = 2z + 4\sqrt{2}z^2 + 8\sqrt{3}z^3 + \cdots$$

Then the theorem says that f is differentiable on the open disc of radius $R = 1/2$, with derivative

$$f'(z) = \sum_{n=0}^{\infty} 2^n n \sqrt{n} z^{n-1} = 2 + 8\sqrt{2}z + 24\sqrt{3}z^2 + \cdots$$

EXERCISE 4.5. Check that the power series for $f'(z)$ in Example 4.20 also has radius of convergence $R = 1/2$.

EXAMPLE 4.21. Consider the power series from EXERCISE 4.3:

$$S(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots .$$

You showed in that exercise that the radius of convergence is $R = +\infty$, so the function $S(z)$ is defined on the entire complex plane $\mathbb{C}$. Using the theorem, we find that the derivative is

$$S'(z) = \sum_{n=0}^{\infty} (-1)^n \frac{(2n+1)z^{2n}}{(2n+1)!} = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!},$$

which is the power series from Example 4.9.

Theorem 4.19 concerns the derivatives of power series, but it immediately implies a result about antiderivatives.

PROPOSITION 4.22. *Suppose that* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *is a power series with radius of convergence* $R > 0$. *Consider the following power series, obtained by antidifferentiating term-by-term using the power rule:*

$$F(z) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} z^{n+1}.$$

The power series $F(z)$ *has the same radius of convergence* R, *and the function* F *provides an antiderivative of* f *on the open disc of radius* R:

$$F'(z) = f(z) \qquad \text{for } |z| < R.$$

PROOF. First suppose that the radius of convergence of $F(z)$ is $S > 0$. Then Theorem 4.19 implies that

$$F'(z) = \sum_{n=0}^{\infty} \frac{(n+1)a_n}{n+1} z^n = \sum_{n=0}^{\infty} a_n z^n = f(z)$$

also has radius of convergence S, which means that $S = R$ and F is an antiderivative of f for $|z| < R$ as claimed.

We now need to investigate the possibility that $S = 0$, which would mean that the power series $F(z)$ converges only for $z = 0$. But this is impossible, for the following reason (we work under the additional assumptions described in Remark 4.6): the radius of convergence R for $f(z)$ is given by $R = 1/L$, where

$$0 \leq L = \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} < +\infty.$$

But then the radius of convergence of $F(z)$ is given by $S = 1/M$, where

$$M = \lim_{n\to\infty} \frac{|a_{n+1}|}{n+1} \cdot \frac{n}{|a_n|} = \lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} \cdot \frac{n}{n+1} = L \cdot 1 = L < +\infty.$$

It follows that $S = 1/M = 1/L > 0$ as claimed. □

EXAMPLE 4.23. Consider again the power series from Example 4.7, with radius of convergence $R = 1/2$:

$$f(z) = \sum_{n=0}^{\infty} 2^n \sqrt{n} z^n = 2z + 4\sqrt{2}z^2 + 8\sqrt{3}z^3 + \cdots$$

The previous proposition says that the following power series $F(z)$ is an antiderivative of $f(z)$ on the open disc of radius $R = 1/2$:

$$F(z) = \sum_{n=0}^{\infty} \frac{2^n \sqrt{n}}{n+1} z^{n+1} = z^2 + \frac{4\sqrt{2}}{3} z^3 + 2\sqrt{3}z^4 + \cdots$$

EXERCISE 4.6. Verify that the antiderivative power series $F(z)$ in Example 4.23 has radius of convergence $R = 1/2$.

In addition to these nice differentiation properties, we may manipulate power series algebraically following the usual rules for polynomials.

PROPOSITION 4.24. *Suppose that* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *has radius of convergence* $R > 0$ *and* $g(z) = \sum_{n=0}^{\infty} b_n z^n$ *has radius of convergence* $S \geq R > 0$. *Then*

(1) *the sum* $f(z)+g(z)$ *may be computed term-by-term, and it has radius of convergence* $\geq R$*:*

$$\begin{aligned} f(z) + g(z) &= \sum_{n=0}^{\infty}(a_n + b_n)z^n \\ &= (a_0 + b_0) + (a_1 + b_1)z + (a_2 + b_2)z^2 + \cdots . \end{aligned}$$

(2) *the product* $f(z)g(z)$ *may be computed following the usual rule for polynomials, and has radius of convergence* $\geq R$*:*

$$\begin{aligned} f(z)g(z) &= \sum_{n=0}^{\infty} c_n z^n \\ &= a_0b_0 + (a_0b_1 + a_1b_0)z + (a_0b_2 + a_1b_1 + a_2b_0)z^2 + \cdots , \end{aligned}$$

where in general the coefficient c_n *is given by*

$$c_n = (a_0b_n + a_1b_{n-1} + \cdots + a_{n-1}b_1 + a_nb_0).$$

REMARK 4.25. As the proof below will show, the formula for the sum $f(z) + g(z)$ holds for any z where both series converge. But the formula for the product $f(z)g(z)$ requires absolute convergence, and hence can only be used safely for z inside the radius of convergence.

PROOF. We will provide a proof of (1), which only requires the limit laws for series. Part (2) is more difficult, and we will omit the proof.

Let c be any complex number for which both power series converge. Letting $s_m(z)$ denote the partial sums of $f(z)$ and $t_m(z)$ denote the partial sums of $g(z)$, we have

$$\begin{aligned}
f(c)+g(c) &= \sum_{n=0}^{\infty} a_n c^n + \sum_{n=0}^{\infty} b_n c^n \\
&= \lim_{m\to\infty} s_m(c) + \lim_{m\to\infty} t_m(c) \\
&= \lim_{m\to\infty} (s_m(c)+t_m(c)) \qquad \text{by Proposition 3.11(a)} \\
&= \lim_{m\to\infty} (a_0 + a_1 c + \cdots + a_m c^m + b_0 + b_1 c + \cdots + b_m c^m) \\
&= \lim_{m\to\infty} ((a_0+b_0) + (a_1+b_1)c + \cdots + (a_m+b_m)c^m) \\
&= \sum_{n=0}^{\infty} (a_n+b_n)c^n .
\end{aligned}$$

Since $S \geq R$, this argument is valid for all c with $|c| < R$, and we have

$$f(z)+g(z) = \sum_{n=0}^{\infty} (a_n+b_n) z^n \qquad \text{for } |z|<R.$$

In particular, the radius of convergence of this power series is $\geq R$. □

Example 4.26. We illustrate the addition and multiplication of power series using the following two series:

$$\begin{aligned}
f(z) &= \sum_{n=0}^{\infty} 2^n z^n = 1 + 2z + 4z^2 + 8z^3 + \cdots \qquad \left(R = \frac{1}{2}\right) \\
g(z) &= \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots \qquad (S = 1).
\end{aligned}$$

The sum $f(z)+g(z)$ also has radius of convergence $R = 1/2$:

$$f(z)+g(z) = \sum_{n=0}^{\infty} (2^n+1) z^n = 2 + 3z + 5z^2 + 9z^3 + \cdots .$$

The product is given by $f(z)g(z) = \sum_{n=0}^{\infty} c_n z^n$, with coefficients

$$\begin{aligned}
c_n &= 2^0 \cdot 1 + 2^1 \cdot 1 + 2^2 \cdot 1 + \cdots + 2^{n-1} \cdot 1 + 2^n \cdot 1 \\
&= 1 + 2 + 2^2 + \cdots + 2^{n-1} + 2^n \\
&= \frac{1-2^{n+1}}{1-2} \\
&= 2^{n+1} - 1.
\end{aligned}$$

Thus, we have the following power series for the product $f(z)g(z)$, with radius of convergence $R = 1/2$:

$$f(z)g(z) = \sum_{n=0}^{\infty}(2^{n+1} - 1)z^n = 1 + 3z + 7z^2 + 15z^3 + \cdots .$$

The results of this section may be summarized as follows: inside its radius of convergence, we are entitled to treat a power series as an "infinite polynomial": it is a differentiable function, and its derivative and antiderivative may be computed term-by-term using the usual power rule. Moreover, power series may be added and multiplied using the same rules as polynomials. In the next section, we exploit these nice features of power series to find series formulas for some familiar functions.

Key points from Section 4.4:

- Term-by-term differentiation and antidifferentiation of power series (Theorem 4.19 and Proposition 4.22)
- Addition and multiplication of power series (Proposition 4.24)

4.5. Power Series Related to Geometric Series

We have made good progress on providing answers to the first three questions about power series listed on page 177, and it is time to move on to question four, repeated here:

Question: Given a complex function $G(z)$, is there a power series formula $G(z) = \sum_{n=0}^{\infty} a_n z^n$, valid on some domain in the complex plane? If so, how can we find the numerical coefficients a_n explicitly?

In the next section we will develop a general procedure to answer this question, applicable to all sorts of complex functions $G(z)$. But in this section, we begin by studying functions that are closely related to

$f(z) = 1/(1-z)$, the function studied in Section 4.1. Since

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \qquad \text{for } |z| < 1,$$

we should be able to use the geometric series as a starting point.

Example 4.27 (Binomial Theorem for Negative Exponents). We begin by taking deriviatives of both sides of the geometric series formula displayed above, using the usual rules of calculus (Proposition 4.18) for the left hand side, and treating the right hand side as an infinite polynomial as explained in the previous section. Taking the first derivative yields the formula

$$\frac{1}{(1-z)^2} = \sum_{n=0}^{\infty} nz^{n-1} = 1 + 2z + 3z^2 + \cdots,$$

valid for all $|z| < 1$. If we continue taking derivatives, we find a whole sequence of formulas:

$$\begin{aligned}
\frac{2}{(1-z)^3} &= \sum_{n=0}^{\infty} n(n-1)z^{n-2} = 2 + 3\cdot 2z + 4\cdot 3z^2 + \cdots \\
\frac{3\cdot 2}{(1-z)^4} &= \sum_{n=0}^{\infty} n(n-1)(n-2)z^{n-3} = 3\cdot 2 + 4\cdot 3\cdot 2z + 5\cdot 4\cdot 3z^2 + \cdots \\
&\vdots \\
\frac{m!}{(1-z)^{m+1}} &= \sum_{n=0}^{\infty} n(n-1)\cdots(n-m+1)z^{n-m}
\end{aligned}$$

After dividing by $m!$, we find that

$$\begin{aligned}
\frac{1}{(1-z)^{m+1}} &= \sum_{n=0}^{\infty} \frac{n(n-1)\cdots(n-m+1)}{m!} z^{n-m} \\
&= \sum_{n=0}^{\infty} \binom{n}{m} z^{n-m} \\
&= 1 + \binom{m+1}{m} z + \binom{m+2}{m} z^2 + \cdots \\
&= \sum_{k=0}^{\infty} \binom{m+k}{m} z^k.
\end{aligned}$$

This equality holds for all $|z| < 1$, and it is called the *binomial theorem for negative integer exponents*, because it generalizes the ordinary binomial theorem:

$$(1+z)^m = \sum_{k=0}^{m} \binom{m}{k} z^k = 1+\binom{m}{1}z+\binom{m}{2}z^2+\cdots+\binom{m}{m-1}z^{m-1}+z^m.$$

So, starting with the geometric series formula for $f(z) = 1/(1-z)$, we have found series formulas for the functions $1/(1-z)^{m+1}$ by taking derivatives and using Theorem 4.19. In the next example, we instead take an antiderivative, using Proposition 4.22.

EXAMPLE 4.28 (Logarithm). We would like to take antiderivatives of both sides of the equation

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + z^3 + \cdots .$$

However, we don't yet know an antiderivative for the complex function $1/(1-z)$, and so we retreat to safety (for the moment) and restrict attention to real values $-1 < x < 1$, where we have the real power series formula

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots .$$

Taking the antiderivatives of both sides now yields

$$-\ln(1-x) + C = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

To find the constant of integration C, simply evaluate both sides at $x = 0$, which shows that $C = 0$:

$$\ln(1-x) = -\sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots$$

We have found a power series formula for the real function $\ln(1-x)$, valid on the open interval $(-1, 1)$. But the same power series formula for the complex variable z converges absolutely on the open unit disc, and there it provides an antiderivative $L(z)$ of the geometric series

$-\sum_{n=0}^{\infty} z^n = -1/(1-z)$:

$$L(z) = -\sum_{n=0}^{\infty} \frac{z^{n+1}}{n+1} = -z - \frac{z^2}{2} - \frac{z^3}{3} - \cdots .$$

We have $L'(z) = -1/(1-z)$ for $|z| < 1$, and we have found a complex antiderivative of the complex function $-1/(1-z)$.

REMARK 4.29. Note that if we evaluate the power series $L(z)$ at a real number $z = x$, then we have

$$L(x) = \ln(1-x) \qquad \text{for } -1 < x < 1.$$

That is, the power series $L(z)$ provides a complex extension of the familiar real function $\ln(1-x)$ to the open unit disc in the complex plane. This example provides a first indication of how power series are relevant to the problem of finding complex extensions of real functions, listed as the 5th question about power series on page 177.

REMARK 4.30. If we replace x by $-x$ in the power series representation of $\ln(1-x)$, we find an alternating series formula for $\ln(1+x)$, valid for $-1 < x < 1$:

$$\ln(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1} = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots .$$

The absolute values of the terms are decreasing, and so we can use the estimation method described in Example 3.38 to obtain approximations to the values of the natural logarithm—see Example 4.57 in Section 4.9.

Moreover, it is tempting to plug in $x = 1$ and obtain the formula

$$\ln(2) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots ,$$

an equality that we wondered about at the end of Section 3.5. While the formula is correct, we have *not* provided a valid derivation. Do you see why?

Sometimes functions need to be manipulated a bit to discover a connection to the geometric series.

EXAMPLE 4.31. Consider the complex function $f(z) = z/(2+3z)$. Observe the following bit of algebraic massaging:

$$f(z) = \frac{z}{2+3z} = z \cdot \frac{1}{2-(-3z)} = \frac{z}{2} \cdot \frac{1}{1-(-\frac{3}{2}z)}.$$

Setting $w = -\frac{3}{2}z$, we see the function $1/(1-w)$ on the right hand side, and we can write this as a geometric series:

$$\begin{aligned} \frac{1}{1-w} &= \sum_{n=0}^{\infty} w^n \\ &= \sum_{n=0}^{\infty} \left(-\frac{3}{2}\right)^n z^n \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{3^n}{2^n} z^n. \end{aligned}$$

This formula is valid for $|w| = \frac{3}{2}|z| < 1$, or $|z| < 2/3$. Finally, we multiply by $z/2$ to obtain a power series formula for the original function $f(z)$, valid on the disc $|z| < 2/3$:

$$\begin{aligned} \frac{z}{2+3z} &= \frac{z}{2} \sum_{n=0}^{\infty} (-1)^n \frac{3^n}{2^n} z^n \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{3^n}{2^{n+1}} z^{n+1} \\ &= \frac{1}{2}z - \frac{3}{4}z^2 + \frac{9}{8}z^3 - \cdots . \end{aligned}$$

Sometimes a connection with the geometric series will only be revealed after taking a derivative or antiderivative.

EXAMPLE 4.32 (Arctangent). Consider $f(x) = \arctan(x)$, the real arctangent function. Recall the formula for the derivative of f from your first calculus course, and observe that it may be massaged into a geometric series:

$$f'(x) = \frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty} (-1)^n x^{2n} = 1 - x^2 + x^4 - x^6 + \cdots .$$

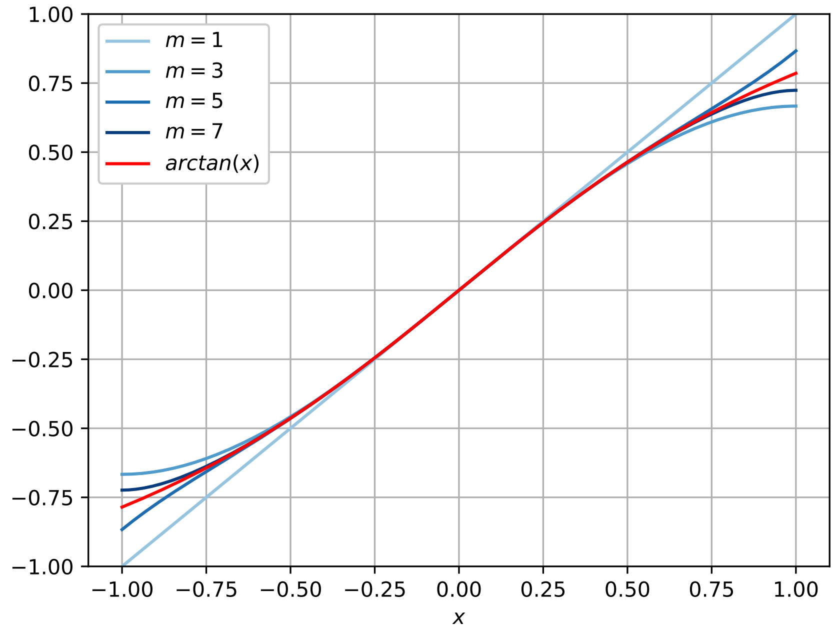


FIGURE 4.6. Plot of the arctangent function on the domain $|x| < 1$ together with the first four distinct partial sums of its series formula.

This formula is valid for $-1 < x < 1$. Now take the antiderivative of both sides to find that

$$f(x) + C = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots .$$

Since $f(0) = \arctan(0) = 0$, the constant of integration $C = 0$, and we find that

$$\arctan(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots ,$$

with radius of convergence $R = 1$. Figure 4.6 shows the first few partial sums as approximations of $\arctan(x)$ on the interval $-1 < x < 1$.

The complex version of this power series also has radius of convergence $R = 1$, so it defines a complex function $A(z)$ on the open unit disc, given by

$$A(z)=\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n+1}}{2n+1}=z-\frac{z^3}{3}+\frac{z^5}{5}-\frac{z^7}{7}+\cdots$$

Note that the function $A(z)$ provides a complex extension of the real arctangent function: for $-1<x<1$, we have $A(x)=\arctan(x)$.

REMARK 4.33. In this section, we have used power series to provide complex extensions of two familiar real functions: the natural logarithm (Example 4.28) and arctangent (Example 4.32). The next exercise asks you to show that these two complex functions are related in an interesting fashion.

EXERCISE 4.7. Let $L(z)$ and $A(z)$ denote the power series from Examples 4.28 and 4.32:

$$L(z)=-\sum_{n=0}^{\infty}\frac{z^{n+1}}{n+1}=-z-\frac{z^2}{2}-\frac{z^3}{3}-\cdots$$

and

$$A(z)=\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n+1}}{2n+1}=z-\frac{z^3}{3}+\frac{z^5}{5}-\frac{z^7}{7}+\cdots$$

Show that for $|z|<1$, we have

$$A(z)=\frac{i}{2}\big(L(iz)-L(-iz)\big)\,.$$

Key points from Section 4.5:

- Binomial Theorem for Negative Exponents (Example 4.27)
- Power series for the logarithm (Example 4.28)
- Manipulating a power series to discover a geometric series (Example 4.31)
- Power series for arctangent (Example 4.32)

4.6. Maclaurin Series

So far, our power series formulas for functions have all been related to the geometric series. But what if we want to find a power series formula for a function like $f(z) = \sqrt{1+z}$ with no clear relationship to the geometric series? To develop a strategy, we employ a common technique of mathematical investigation: we assume that we already have what we are looking for, and then we try to learn something about it. So suppose that $f(z)$ is a function, and that there is a power series formula for $f(z)$, valid inside some radius of convergence $R > 0$:

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots .$$

First note that by Theorem 4.19, the power series on the right hand side can be differentiated as many times as we want on the open disc of radius R, hence the same must be true for the function $f(z)$. So we have learned something:

- If the function $f(z)$ has a power series formula on a disc of positive radius, then $f(z)$ is infinitely differentiable on that disc.

Moreover, using term-by-term differentiation, we see that the coefficients a_n are completely determined by the derivatives of f at $z = 0$:

$$\begin{aligned}
&f(z) = a_0 + a_1 z + \cdots , && \text{so } f(0) = a_0\\
&f'(z) = a_1 + 2a_2 z + \cdots , && \text{so } f'(0) = a_1\\
&f''(z) = 2a_2 + 3\cdot 2a_3 z + \cdots , && \text{so } f''(0) = 2a_2\\
&f'''(z) = 3\cdot 2a_3 + 4\cdot 3\cdot 2a_4 z + \cdots , && \text{so } f'''(0) = 3\cdot 2a_3\\
&\vdots && \vdots\\
&f^{(n)}(z) = n!a_n + (n+1)!a_{n+1} z + \cdots , && \text{so } f^{(n)}(0) = n!a_n\\
&\vdots && \vdots
\end{aligned}$$

We turn this computation into a definition.

DEFINITION 4.34 (Maclaurin Series). Suppose that $f(z)$ is a function defined and infinitely differentiable near $z = 0$. The *Maclaurin* *series of* f is the power series with coefficients $a_n = f^{(n)}(0)/n!$:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n.$$

Remark 4.35. We emphasize that the Maclaurin coefficients a_n are *constants* obtained by *evaluating* the derivatives of f at $z = 0$ and then dividing by factorials. A common mistake is to omit the evaluation at zero, which leads to a more complicated expression that is *not* the Maclaurin series. For instance, consider the exponential function $f(x) = e^x$. Then $f^{(n)}(x) = e^x$ for all $n \geq 0$, and so the Maclaurin coefficients a_n are given by

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{e^0}{n!} = \frac{1}{n!}.$$

The Maclaurin series of e^x is therefore

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots .$$

Note how different this is from the following series, in which we forget to evaluate the derivatives at $x = 0$:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} x^n = \sum_{n=0}^{\infty} \frac{e^x}{n!} x^n = e^x + e^x x + e^x \frac{x^2}{2!} + e^x \frac{x^3}{3!} + \cdots .$$

In fact, this series is not even a *power* series, and it is certainly not the Maclaurin series of e^x.

The discussion preceding Definition 4.34 establishes the following:

- If a function $f(z)$ has a power series formula on a disc of positive radius, then that power series must be the Maclaurin series of f.

To make sense of the coefficients in the Maclaurin series, consider the mth partial sum:

$$s_m(z) = f(0) + f'(0)z + \frac{f''(0)}{2!} z^2 + \cdots + \frac{f^{(m)}(0)}{m!} z^m.$$

This is a polynomial of degree m that has the same first m derivatives at zero as the function f:

$$s_m(0) = f(0), \quad s'_m(0) = f'(0), \quad s''_m(0) = f''(0), \cdots, s_m^{(m)}(0) = f^{(m)}(0).$$

It follows that $s_m(z)$ is the best mth degree polynomial approximation to f at $z = 0$, in the sense discussed on page 177.

We record our findings in the following strategy:

Strategy: In order to find a power series formula for a function $f(z)$:

(1) Compute all derivatives $f^{(n)}(0)$ for $n \geq 0$. (If f is not infinitely differentiable, then it has no power series formula.)

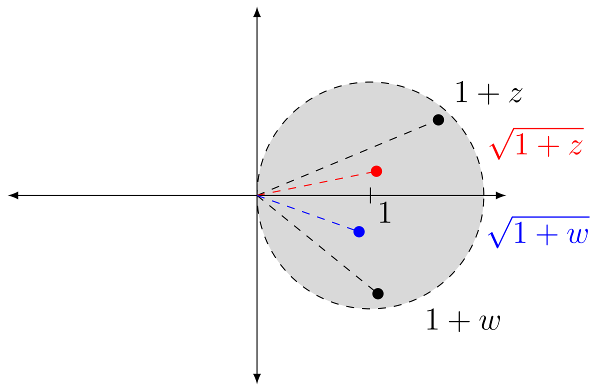
(2) Write down the Maclaurin series for f and find its radius of convergence R:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n.$$

(3) Try to prove that the Maclaurin series is equal to the function $f(z)$ for $|z| < R$:

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n \qquad \text{for } |z| < R.$$

In the next example, we will employ this strategy for the function $f(z) = \sqrt{1+z}$. Before we get started, let's be sure we understand the meaning of this complex function. We are interested in values of z near zero, which means that $1+z$ is near 1. In particular, if $|z| < 1$, then $1+z$ is contained in the open disc of radius 1 centered at 1 (see picture below). This disc does not include zero, so every number $1+z$ in the disc has two distinct square roots. Which one of these square roots are we referencing with the function $f(z) = \sqrt{1+z}$? We make the following choice, which matches the convention from Example 1.19 in Chapter 1: each point $1+z$ in the disc has polar coordinates of the form (r, θ) with $-\pi/2 < \theta < \pi/2$. Negative arguments correspond to points in the bottom of the disc, positive arguments to points in the top. Then $f(z) = \sqrt{1+z}$ is the square root with polar coordinates $(\sqrt{r}, \theta/2)$.



EXAMPLE 4.36. Let's compute the Maclaurin series of the complex function $f(z) = \sqrt{1+z}$. Of course, we don't yet actually know that $f(z)$ is complex differentiable, nor that the usual power rule applies if it is differentiable. So for the moment, let's restrict attention to real values x, so we have $f(x) = \sqrt{1+x}$. Here is the sequence of derivatives:

$$\begin{aligned} f(x) &= (1+x)^{\frac{1}{2}} \\ f'(x) &= \frac{1}{2}(1+x)^{-\frac{1}{2}} \\ f''(x) &= -\frac{1}{4}(1+x)^{-\frac{3}{2}} \\ f'''(x) &= \frac{3}{8}(1+x)^{-\frac{5}{2}} \\ &\vdots \\ f^{(n)}(x) &= (-1)^{n+1}\frac{(2n-3)!!}{2^n}(1+x)^{-\frac{2n-1}{2}} \quad \text{for } n \geq 2. \end{aligned}$$

(Here, the *double factorial* $(2n-3)!!$ indicates the product of all odd numbers between 1 and $2n-3$.) Evaluating at $x = 0$ and dividing by $n!$ yields the coefficients a_n of the Maclaurin series:

$$a_0 = 1,\ a_1 = \frac{1}{2},\ a_2 = -\frac{1}{8},\ a_3 = \frac{1}{16}, \cdots, a_n = (-1)^{n+1}\frac{(2n-3)!!}{2^n n!}.$$

So the real Maclaurin series for $f(x) = \sqrt{1+x}$ is

$$1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \cdots = 1 + \frac{x}{2} + \sum_{n=2}^{\infty} (-1)^{n+1} \frac{(2n-3)!!}{2^n n!} x^n.$$

Here is the corresponding complex series:

$$1 + \frac{z}{2} + \sum_{n=2}^{\infty} (-1)^{n+1} \frac{(2n-3)!!}{2^n n!} z^n.$$

To find the radius of convergence, we investigate the consecutive ratios:

$$\frac{|a_{n+1}|}{|a_n|} = \frac{(2n-1)!!}{2^{n+1}(n+1)!} \cdot \frac{2^n n!}{(2n-3)!!} = \frac{2n-1}{2(n+1)}.$$

Taking the limit yields

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \frac{2n-1}{2(n+1)} = \lim_{n\to\infty} \frac{1 - \frac{1}{2n}}{1 + \frac{1}{n}} = 1 = L,$$

which implies that the radius of convergence is $R = 1/L = 1$. So the complex Maclaurin series defines *some* function on the open unit disc, but how do we know that the function is $f(z) = \sqrt{1+z}$?

Figure 4.7 shows pictures of some partial sums of the real series, which suggest that they are indeed converging to $\sqrt{1+x}$. But how can we know for sure, and what about the complex series? We could try to check explicitly that the square of the complex series is equal to $1+z$. This seems like a daunting computation, so we will just square each of the first few partial sums:

$$\begin{aligned}
\left(1 + \frac{z}{2}\right)^2 &= 1 + z + \frac{z^2}{4} \\
\left(1 + \frac{z}{2} - \frac{z^2}{8}\right)^2 &= 1 + z - \frac{z^3}{8} + \frac{z^4}{64} \\
\left(1 + \frac{z}{2} - \frac{z^2}{8} + \frac{z^3}{16}\right)^2 &= 1 + z + \frac{5z^4}{64} - \frac{z^5}{64} + \frac{z^6}{256}
\end{aligned}$$

Note that, at least in these initial cases, we have

$$s_m(z)^2 = 1 + z + \text{terms of degree greater than } m.$$

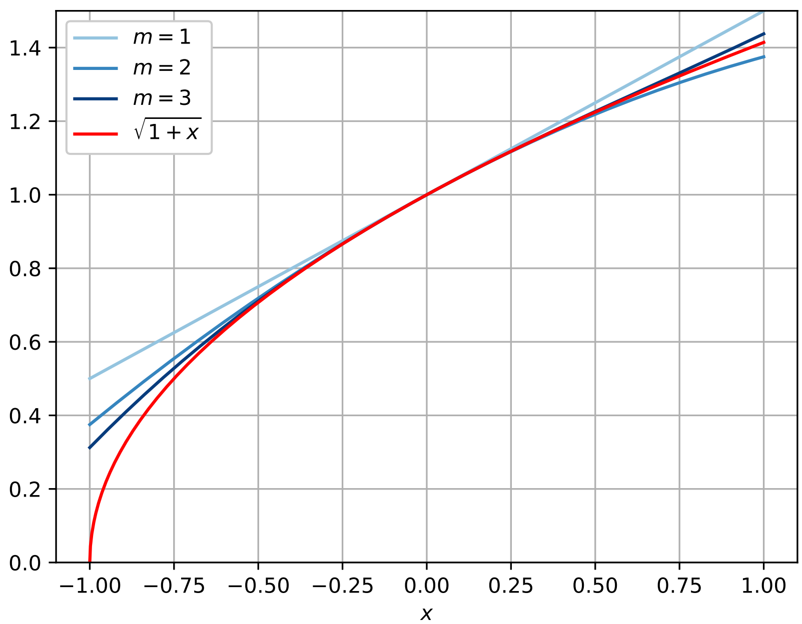


FIGURE 4.7. Three polynomial partial sums $s_m(x)$ for the Maclaurin series of $\sqrt{1+x}$.

Assuming that this pattern continues (it does), the full Maclaurin series will have square equal to $1+z$. Of course, this method of verification is very special to the function $\sqrt{1+z}$, and will not work in general to settle the question of whether or not a Maclaurin series represents its function. In Example 4.60 in the optional Section 4.9, we will study the functions $(1+z)^p$ for real exponents p, and prove that they are represented by their Maclaurin series; the square root function of this example is the case $p = 1/2$.

REMARK 4.37. There are examples of infinitely differentiable real functions $f(x)$ that are not represented by their Maclaurin series on any interval. The standard example is

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

As suggested by Figure 4.8, this function is so flat at the origin that *all* of its derivatives are zero: $f^{(n)}(0) = 0$ for all $n \geq 0$. It follows that its Maclaurin series is the constant function 0, which does not agree with the nonconstant function $f(x)$ except at the single point $x = 0$.

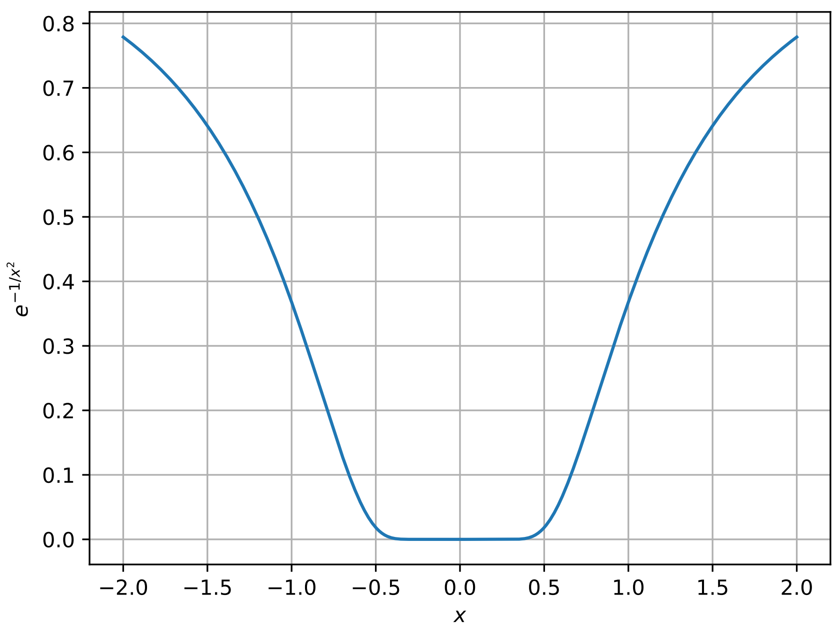


FIGURE 4.8. Graph of the function $f(x) = e^{-\frac{1}{x^2}}$.

We now state an amazing theorem which implies that the phenomenon illustrated in the previous example cannot happen for complex functions:

THEOREM 4.38. *If $f(z)$ is complex differentiable in an open disc centered at zero, then it is actually infinitely differentiable and represented by its Maclaurin series on that disc.*

PROOF. Take MATH 535: Complex Analysis. □

COROLLARY 4.39. *Suppose that $f(z)$ and $g(z)$ are both complex differentiable in an open disc centered at zero, and that $f(x) = g(x)$*

for all real x in an open interval containing zero. Then $f(z) = g(z)$ for all z in the open disc.

Proof. Consider the difference $h(z) = f(z) - g(z)$, which is also complex differentiable on the open disc, hence infinitely differentiable by the previous theorem. Moreover, $h(x) = 0$ for all x in some open interval containing zero. Since we can compute $h'(x)$ by approaching x along the real axis, it follows that $h'(x) = 0$ for all x in the open interval. Repeating this argument, we find that the nth derivative $h^{(n)}(x) = 0$ for all x near zero. In particular, $h^{(n)}(0) = 0$ for all $n \geq 0$, which implies that the Maclaurin series of h is 0. But this series represents h on the open disc by the theorem, which means that $h = 0$, or $f(z) = g(z)$ for all z in the open disc. □

This corollary forms the basis of our approach to the problem of finding complex extensions of real functions (question (5) from page 177). Namely: suppose that the real function $f(x)$ is defined near zero, and suppose that we successfully employ our Maclaurin series strategy to find a real power series formula for $f(x)$:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \qquad \text{for } |x| < R.$$

Then the corresponding complex power series has the same radius of convergence R, and so it defines a complex extension $F(z)$ of the real function $f(x)$ to a disc in the complex plane:

$$\begin{aligned} F(z) &= \sum_{n=0}^{\infty} a_n z^n && \text{for } |z| < R, \\ F(x) &= f(x) && \text{for } -R < x < R. \end{aligned}$$

Moreover, the corollary implies that this is the *only* possible extension that is complex differentiable.

Note that this is the method we employed to find complex extensions of the natural logarithm and arctangent functions in Examples 4.28 and 4.32. In the next section, we will use this strategy to find complex extensions of the real functions $\cos(x), \sin(x)$, and e^x.

> Key points from Section 4.6:
> - Definition of Maclaurin series (Definition 4.34)
> - Strategy for finding power series formulas (page 206)

4.7. The Complex Exponential, Sine, and Cosine

We now employ the ideas of the previous section to derive power series formulas for the real functions $e^x, \cos(x)$, and $\sin(x)$, together with extensions of these functions to the complex plane. For each function $f(x)$, we postpone the key step of the argument: showing that the real Maclaurin series actually converges to the real function $f(x)$. This step will be explained carefully in the optional Section 4.8.

Example 4.40 (Exponential Function). Consider the real exponential function $f(x) = e^x$, which is defined for all real x. For all $n \geq 0$ we have $f^{(n)}(x) = e^x$, so the Maclaurin coefficients are

$$a_n = \frac{f^{(n)}(0)}{n!} = \frac{e^0}{n!} = \frac{1}{n!},$$

and the Maclaurin series is

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots .$$

In Example 3.43, we showed that this series converges for all values of x, and hence has radius of convergence $R = +\infty$. Figure 4.9 shows the first few partial sums as approximations of $f(x) = e^x$, which suggests (but does not prove!) that the series converges to e^x.

The corresponding complex series defines a function on the entire complex plane, called the *complex exponential function*:

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Note that we are introducing a new symbol on the left hand side, $\exp(z)$, which is simply our name for the function defined by the power

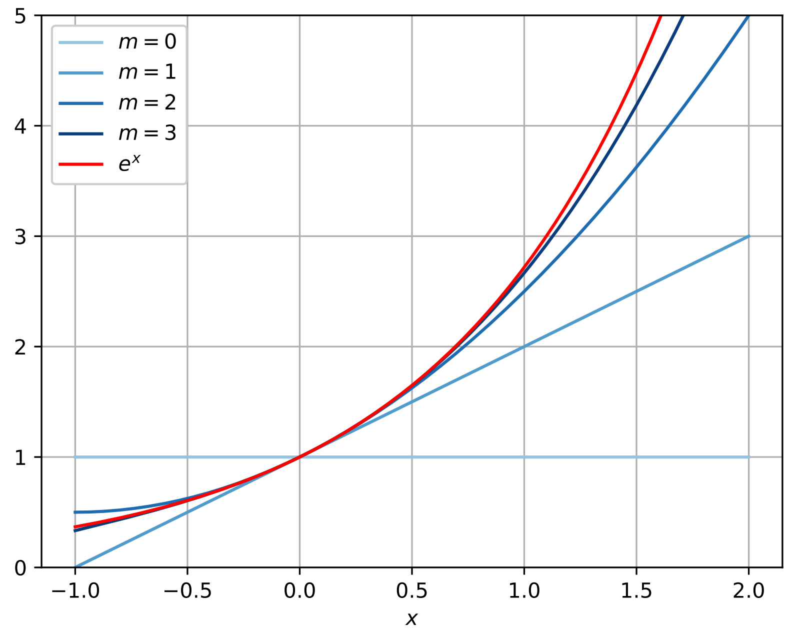


FIGURE 4.9. Plot of the exponential function together with the first four partial sums of its Maclaurin series.

series on the right hand side. Once we prove (in the optional Section 4.8) that the real Maclaurin series converges to e^x, then we will know that $\exp\colon \mathbb{C} \to \mathbb{C}$ is the unique complex differentiable extension of e^x to the complex plane, thus justifying its name.

EXAMPLE 4.41 (Sine). Now consider the sine function $f(x) = \sin(x)$, also defined for all real x. We start by computing some derivatives:

$$\begin{aligned} f(x) &= \sin(x) \\ f'(x) &= \cos(x) \\ f''(x) &= -\sin(x) \\ f^{(4)}(x) &= -\cos(x) \\ f^{(5)}(x) &= \sin(x) \\ &\vdots \end{aligned}$$

We see the pattern: even derivatives are $\pm\sin(x)$ with the signs alternating, while odd derivatives are $\pm\cos(x)$ with the signs alternating. Moreover, even integers may be written as $n = 2k$ for some k, while odd integers may be written as $n = 2k+1$ for some k. Putting this all together, we find that

$$f^{(n)}(x) = \begin{cases} (-1)^k \sin(x) & \text{if } n = 2k \text{ is even} \\ (-1)^k \cos(x) & \text{if } n = 2k+1 \text{ is odd.} \end{cases}$$

Evaluating at $x = 0$ and dividing by $n!$ yields the Maclaurin coefficients:

$$a_n = \begin{cases} 0 & \text{if } n = 2k \text{ is even} \\ (-1)^k/(2k+1)! & \text{if } n = 2k+1 \text{ is odd.} \end{cases}$$

The Maclaurin series is

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

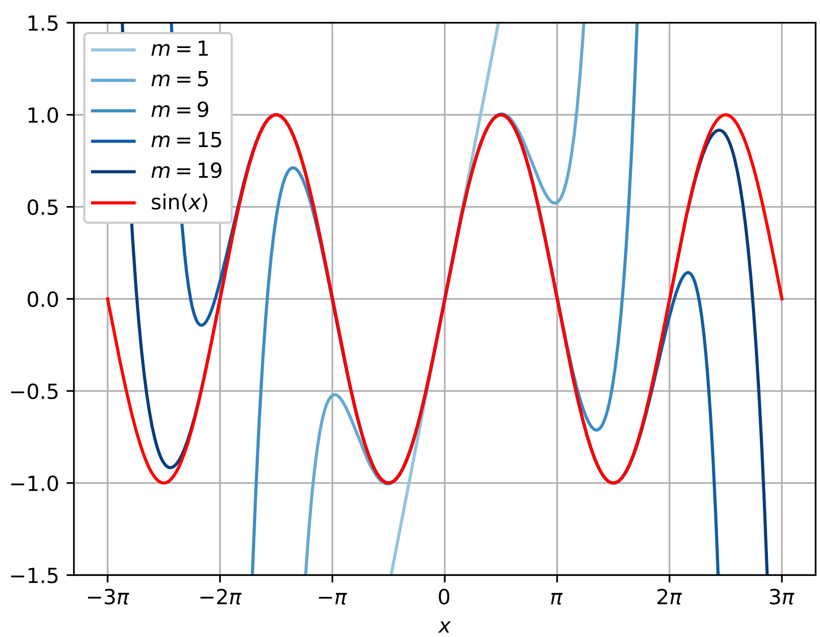


FIGURE 4.10. Plot of the sine function together with some partial sums of its Maclaurin series.

In EXERCISE 4.3, you showed that this series has radius of convergence $R = +\infty$. The corresponding complex series thus defines a complex differentiable function on $\mathbb{C}$, which we call the *complex sine function*:

$$\sin(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}.$$

Once again, the symbol $\sin(z)$ on the left hand side is simply our name for the function defined by the power series on the right hand side. This name will be justified once we prove (in the optional Section 4.8) that the real Maclaurin series converges to the real sine function $\sin(x)$. Figure 4.10 shows a few partial sums as approximations of $\sin(x)$, which provide some visual evidence for this assertion.

EXERCISE 4.8. (Cosine) Compute the Maclaurin series for $\cos(x)$, and show that the complex version matches the series from Example 4.9, with radius of convergence $R = +\infty$:

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

So how can we prove that a real function $f(x)$ is represented by its Maclaurin series? Well, let's see what this really means: if $f(x)$ is equal to its series for $|x| < R$, then we have

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = \lim_{m \to \infty} s_m(x),$$

the limit of the polynomial partial sums $s_m(x)$. Subtracting, we find that

$$0 = \lim_{m \to \infty} \big(f(x) - s_m(x)\big) = \lim_{m \to \infty} R_m(x) \qquad \text{for } |x| < R,$$

where $R_m(x) = f(x) - s_m(x)$ is the *mth Maclaurin remainder of* $f(x)$.

So: in order to show that $f(x)$ is represented by its Maclaurin series, we need to show that for $|x| < R$, the sequence of remainders $R_m(x)$ converges to zero. For this, it would be helpful to have a result that tells us something about the size of the remainders $R_m(x)$ if we know something about the function $f(x)$ and all of its derivatives near 0. *Taylor's Theorem* in the next section provides just what we need.

Because the proofs require Taylor's Theorem and are a bit technical, we postpone until the next optional section the arguments that $e^x, \sin(x)$, and $\cos(x)$ are each represented by their Maclaurin series. For the remainder of this section, we explore some remarkable relationships between the complex extensions $\exp(z), \sin(z)$, and $\cos(z)$. The next exercise is essential—be sure to complete it before you move on.

EXERCISE 4.9. (Euler's Formula) Recall the power series defining the complex exponential, sine, and cosine as functions on the complex plane:

$$\begin{aligned}
\exp(z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \\
\sin(z) &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots \\
\cos(z) &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots
\end{aligned}$$

Using these power series, show that for all complex numbers z,

$$\exp(iz) = \cos(z) + i\sin(z).$$

This result is likely to surprise you, since your prior experience with the exponential and trigonometric functions may not suggest much of a connection between them. But as Jacques Hadamard observed in 1945: "It has been written that the shortest and best way between two truths of the real domain often passes through the imaginary one."[1]

[1] *An Essay on the Psychology of Invention in the Mathematical Field* (Princeton U. Press, 1945, p. 123)

Euler's formula is often presented for real numbers $z = y$, in which case it says that

$$\exp(iy) = \cos(y) + i\sin(y).$$

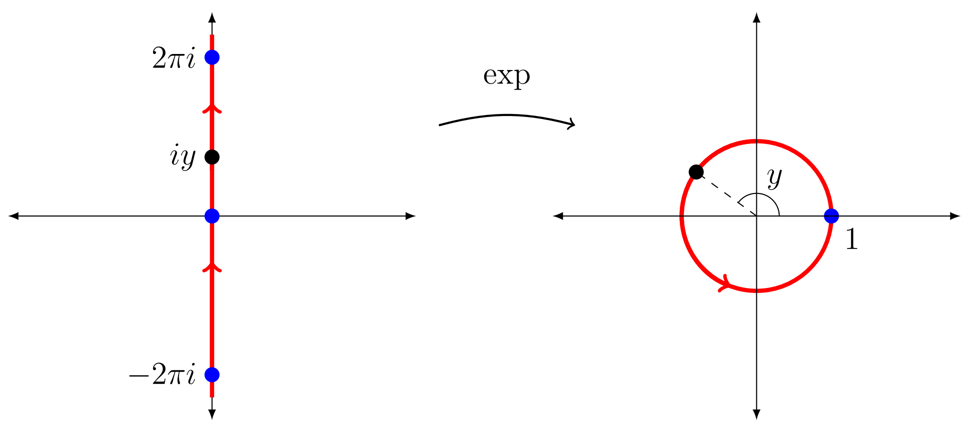
Note that the magnitude of these numbers is always 1:

$$\big|\exp(iy)\big|^2 = \cos^2(y) + \sin^2(y) = 1.$$

Moreover, since cosine and sine each have period 2π, we find that

$$\exp(2\pi i n) = \cos(2\pi n) + i\sin(2\pi n) = 1 \qquad \text{for all integers } n.$$

These formulas have the following geometric interpretation: the complex exponential function takes the imaginary axis and wraps it infinitely many times counterclockwise around the unit circle, making one revolution every time y increases by 2π. That is, $\exp(iy)$ is the point on the unit circle with argument $\theta = y$ (see picture below).



In order to understand more about the geometry of the complex exponential, we need to establish the complex generalization of the familiar law of real exponents $e^{x+y} = e^x e^y$.

PROPOSITION 4.42. *For any two complex numbers z and w, we have*

$$\exp(z + w) = \exp(z)\exp(w).$$

PROOF. This follows from the multiplication of power series as infinite polynomials, Proposition 4.24.

$$\begin{aligned}
\exp(z)\exp(w) &= (1 + z + \frac{z^2}{2} + \frac{z^3}{3!} + \cdots)(1 + w + \frac{w^2}{2} + \frac{w^3}{3!} + \cdots) \\
&= 1 + (z + w) + (\frac{z^2}{2} + zw + \frac{w^2}{2}) \\
&\qquad + (\frac{z^3}{3!} + \frac{z^2 w}{2} + \frac{zw^2}{2} + \frac{w^3}{3!}) + \cdots \\
&= 1 + (z + w) + \frac{z^2 + 2zw + w^2}{2} \\
&\qquad + \frac{z^3 + 3z^2 w + 3zw^2 + w^3}{3!} + \cdots \\
&= 1 + (z + w) + \frac{(z + w)^2}{2} + \frac{(z + w)^3}{3!} + \cdots \\
&= \exp(z + w).
\end{aligned}$$

To see that the suggested pattern really does continue, note that for any $n \geq 0$, the part of total degree n on the right hand side will form a sum with $n + 1$ terms:

$$\begin{aligned}
\sum_{k=0}^{n} \frac{z^{n-k} w^k}{(n-k)!k!} &= \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} \cdot \frac{z^{n-k} w^k}{n!} \\
&= \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^{n-k} w^k \\
&= \frac{(z + w)^n}{n!}.
\end{aligned}$$

by the binomial theorem. □

We apply this result to $x + iy$ and find that

$$\exp(x + iy) = \exp(x)\exp(iy) = e^x(\cos(y) + i\sin(y)),$$

where we have used Euler's formula in the final step, together with the (as yet unproven) assertion that $\exp(x) = e^x$. You may recognize this from long ago in Chapter 1 as the function from Example 1.20. You should review that example carefully now to recall the geometry of the complex exponential function.

EXAMPLE 4.43 (de Moivre's Formula). A nice application of Euler's formula $\exp(i\theta) = \cos(\theta) + i\sin(\theta)$ is to produce trigonometric identities: raise each side of the formula to the nth power and use Proposition 4.42:

$$\exp(in\theta) = (\exp(i\theta))^n = (\cos(\theta) + i\sin(\theta))^n.$$

But Euler's formula says that $\exp(in\theta) = \cos(n\theta) + i\sin(n\theta)$, which yields *de Moivre's formula*:

$$\cos(n\theta) + i\sin(n\theta) = (\cos(\theta) + i\sin(\theta))^n.$$

Expanding the right hand side using the binomial theorem and collecting the real and imaginary parts yields the *multiple angle identities*:

$n = 2$:

$$\begin{aligned}\cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta)\\ \sin(2\theta) &= 2\cos(\theta)\sin(\theta)\end{aligned}$$

$n = 3$:

$$\begin{aligned}\cos(3\theta) &= \cos^3(\theta) - 3\cos(\theta)\sin^2(\theta)\\ \sin(3\theta) &= 3\cos^2(\theta)\sin(\theta) - \sin^3(\theta)\end{aligned}$$

$n = 4$:

$$\begin{aligned}\cos(4\theta) &= \cos^4(\theta) - 6\cos^2(\theta)\sin^2(\theta) + \sin^4(\theta)\\ \sin(4\theta) &= 4\cos^3(\theta)\sin(\theta) - 4\cos(\theta)\sin^3(\theta),\end{aligned}$$

etc.

EXERCISE 4.10. Use de Moivre's formula to find identities for $\cos(5\theta)$ and $\sin(5\theta)$ in terms of $\cos(\theta)$ and $\sin(\theta)$.

At the beginning of the course we saw two advertisements for the superiority of the complex numbers over the real numbers:

(1) The Fundamental Theorem of Algebra (Theorem 1.16): every nonconstant complex polynomial has a complex root. The

analogous statement is false for the real numbers, as there are real quadratic polynomials (e.g. x^2+1) with no real roots.

(2) Complex Dynamics (Section 2.1): our experiments with iterating complex quadratic polynomials led to a striking fractal set in the complex plane. The corresponding set of real numbers is just the closed interval $[-2, 1/4]$.

Our work in this section serves as a third advertisement: whereas the real exponential function e^x seems to have no connection with the trigonometric functions $\cos(x)$ and $\sin(x)$, the complex exponential $\exp(z)$ unifies all three into a single function displaying beautiful geometry and providing a coherent explanation for some familiar trigonometric identities. See the optional Section 5.4 of the next chapter for the connection of the exponential function to a different form of trigonometry based on the hyperbola rather than the circle.

Given how often you have encountered the real functions $e^x, \sin(x)$, and $\cos(x)$, it should come as no surprise that the complex versions are also ubiquitous in mathematics, physics, and engineering. As examples, in Sections 5.2 and 5.3 of the next chapter, we explain how these functions appear in the study of differential equations and oscillating physical systems.

Key points from Section 4.7:

- Maclaurin series for $\exp(z), \sin(z)$, and $\cos(z)$. (Example 4.40, 4.41 and Exercise 4.8)
- Euler's formula (Exercise 4.9)
- Multiplication law for $\exp(z)$ (Proposition 4.42)

4.8. Optional: Taylor Series and Taylor's Theorem

All of our work so far has focused on the behavior of functions near $z = 0$. But we may wish to investigate the behavior of functions near an arbitrary point $z = c$ in the complex plane. In the following example, we illustrate by reference to the function $f(z) = 1/(1-z)$.

EXAMPLE 4.44. Suppose we want to study the complex function $f(z) = 1/(1-z)$ near the point $z = 2$. Note that this is outside the domain of convergence of the geometric series $\sum_{n=0}^{\infty} z^n$, which only represents $f(z)$ on the unit disc centered at 0. But we can use a different geometric series if we massage the formula a bit:

$$\begin{aligned}
\frac{1}{1-z} &= \frac{1}{-1-(z-2)} \\
&= \frac{-1}{1+(z-2)} \\
&= \frac{-1}{1-(-(z-2))} \\
&= -\sum_{n=0}^{\infty}(-1)^n(z-2)^n \\
&= -1+(z-2)-(z-2)^2+(z-2)^3-\cdots
\end{aligned}$$

This series formula is valid for $|z-2| < 1$, which defines the open unit disc centered at 2.

We now extend some our earlier concepts about power series to this more general setting.

DEFINITION 4.45. Suppose that c is a fixed complex number. Then *the power series centered at c with coefficients a_n* is the following infinite series, viewed as a function of the complex variable z:

$$F(z) = \sum_{n=0}^{\infty} a_n(z-c)^n = a_0 + a_1(z-c) + a_2(z-c)^2 + \cdots$$

All of the results from Section 4.4 continue to hold for these more general power series. In particular, every power series centered at c has a radius of converge $R \geq 0$ or $R = +\infty$. Inside the disc of radius R centered at c, the series converges absolutely to a complex differentiable function that may be differentiated and antidifferentiated term-by-term. Moreover, we may perform addition and multiplication for power series centered at c just as if they were "infinite polynomials."

Note that if the center c and the coefficients a_n are real numbers, then we get a real power series centered at c, which converges absolutely to an infinitely differentiable function on the open interval $(c-R, c+R)$.

EXAMPLE 4.46. Here is a power series centered at $c = 2$:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n \cdot 2^n}(z-2)^n = \frac{1}{2}(z-2) - \frac{1}{8}(z-2)^2 + \frac{1}{24}(z-2)^3 - \cdots$$

To find the radius of convergence, we use the ratio test:

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \frac{n \cdot 2^n}{(n+1)\cdot 2^{n+1}} = \lim_{n\to\infty} \frac{1}{2} \cdot \frac{n}{n+1} = \frac{1}{2} = L.$$

The ratio test guarantees absolute convergence for $|z-2| < 1/L = 2$, so the radius of convergence is $R = 2$.

DEFINITION 4.47 (Taylor Series). Suppose that $f(z)$ is a complex function defined and infinitely differentiable near $z = c$.

(1) The *Taylor series of f at c* is the power series centered at c with coefficients $a_n = f^{(n)}(c)/n!$:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(z-c)^n.$$

(2) The mth partial sum of the Taylor series is called the *mth Taylor polynomial for f at c*:

$$T_m(z) = f(c) + f'(c)(z-c) + \frac{f''(c)}{2}(z-c)^2 + \cdots + \frac{f^{(m)}(c)}{m!}(z-c)^m.$$

As before, the polynomial $T_m(z)$ is the best mth degree approximation to $f(z)$ at $z = c$, in the sense that $T_m^{(k)}(c) = f^{(k)}(c)$ for $0 \le k \le m$.

(3) The *mth Taylor remainder of f at c* is the difference between the function $f(z)$ and the mth Taylor polynomial:

$$R_m(z) = f(z) - T_m(z).$$

If $f(x)$ is instead a real function defined and infinitely differentiable near the real number $x = c$, then these same definitions yield the real Taylor series, polynomials, and remainders for f at c.

EXAMPLE 4.48. Consider the real function $f(x) = \ln(x)$ near the point $c = 2$. Here is the sequence of derivatives:

$$\begin{aligned} f'(x) &= \frac{1}{x} \\ f''(x) &= -\frac{1}{x^2} \\ f'''(x) &= \frac{2}{x^3} \\ f^{(4)}(x) &= -\frac{3 \cdot 2}{x^4} \\ &\vdots \\ f^{(n)}(x) &= (-1)^{n+1} \frac{(n-1)!}{x^n}. \end{aligned}$$

Evaluating at $x = 2$ and dividing by $n!$ yields the the Taylor coefficients a_n:

$$a_0 = f(2) = \ln(2), \qquad a_n = \frac{f^{(n)}(2)}{n!} = (-1)^{n+1} \frac{(n-1)!}{n! \cdot 2^n} = \frac{(-1)^{n+1}}{n \cdot 2^n}.$$

Thus, the Taylor series of f at $c = 2$ is basically the series from Example 4.46:

$$\ln(2) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n \cdot 2^n} (x-2)^n.$$

We are interested in showing that particular real functions $f(x)$ are represented by their Taylor series near $x = c$. As discussed on page 215 for Maclaurin series, we need to show that the Taylor remainders $R_m(x)$ converge to zero as $m \to \infty$ for all x near c. We repeat the logic here: to say that

$$f(x) = \sum_{n=0}^{\infty} a_n (x-c)^n = \lim_{m \to \infty} T_m(x) \qquad (|x-c| < R)$$

is equivalent to saying that

$$\lim_{m \to \infty} R_m(x) = \lim_{m \to \infty} (f(x) - T_m(x)) = f(x) - \lim_{m \to \infty} T_m(x) = 0 \quad (|x-c| < R).$$

The following important theorem provides just the tool we need, namely information about the remainders $R_m(x)$ in terms of the derivatives of f near $x = c$.

THEOREM 4.49 (Taylor's Theorem). *Suppose that the real function $f(x)$ is $(m+1)$-times continuously differentiable on the open interval $(c-R, c+R)$. Then for each x with $|x-c|<R$, there exists a point $\tilde{x}$ between c and x such that*

$$R_m(x) = f^{(m+1)}(\tilde{x})\frac{(x-c)^{m+1}}{(m+1)!}$$

(Note that the point $\tilde{x}$ depends on x.)

Before proving Taylor's Theorem, we show how to use it.

EXAMPLE 4.50 (Exponential). Consider the real exponential function $f(x) = e^x$, and the point $c = 0$. Then the Taylor series at $c = 0$ is the Maclaurin series computed in Example 4.40:

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

We will use Taylor's Theorem to show that the remainders $R_m(x)$ converge to zero as $m \to \infty$ for all x. Note that $f(x) = e^x$ does satisfy the necessary hypotheses, as f is infinitely differentiable on $\mathbb{R}$, with $f^{(n)}(x) = e^x$ for all $n \geq 0$. So fix x and m, and look at the conclusion of Taylor's Theorem:

$$R_m(x) = e^{\tilde{x}}\frac{x^{m+1}}{(m+1)!}.$$

Here, $\tilde{x}$ is some unknown real number with $|\tilde{x}| \leq |x|$. In particular, $e^{\tilde{x}} \leq e^{|x|}$, and so we have

$$|R_m(x)| \leq e^{|x|}\frac{|x|^{m+1}}{(m+1)!}.$$

By Lemma 3.44 (with $c = |x|$) the right hand side converges to zero as $m \to \infty$. Hence, we have shown that $\lim_{m\to\infty} R_m(x) = 0$. It follows, at long last, that the real exponential function is represented by its Maclaurin series for all real x:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

EXAMPLE 4.51 (Sine). Consider the sine function $\sin(x)$. Again, fix x and m and examine the conclusion of Taylor's Theorem:

$$R_m(x) = \begin{cases} (-1)^k \sin(\tilde{x}) \frac{x^{2k}}{(2k)!} & \text{if } m = 2k-1 \text{ is odd} \\ (-1)^k \cos(\tilde{x}) \frac{x^{2k+1}}{(2k+1)!} & \text{if } m = 2k \text{ is even.} \end{cases}$$

Here, $\tilde{x}$ is an unknown real number with $|\tilde{x}| \leq |x|$. But both sine and cosine are bounded by 1 in absolute value, so we find that

$$|R_m(x)| \leq \begin{cases} \frac{|x|^{2k}}{(2k)!} & \text{if } m = 2k-1 \text{ is odd} \\ \frac{|x|^{2k+1}}{(2k+1)!} & \text{if } m = 2k \text{ is even.} \end{cases}$$

Both of the sequences on the right hand side converge to zero by Lemma 3.44. It follows that $\sin(x)$ is represented by its Maclaurin series for all real x:

$$\sin(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

EXERCISE 4.11. (Cosine) Use Taylor's Theorem to show that $\cos(x)$ is represented by its Maclaurin series for all real x.

EXAMPLE 4.52 (Logarithm). As a final example, we use Taylor's Theorem to show that $f(x) = \ln(x)$ is represented by its Taylor series near $c = 2$. We computed the Taylor series in Example 4.48:

$$\ln(2) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n \cdot 2^n} (x-2)^n.$$

This power series has radius of converge $R = 2$, but we will only show that it converges to $\ln(x)$ for $|x - 2| < 1$.

So fix m and a point x with $|x-2|<1$. Then by Taylor's Theorem there is a point $\tilde{x}$ between 2 and x such that

$$\begin{aligned} \left|R_m(\tilde{x})\right| &= \left|f^{(m+1)}(\tilde{x})\frac{(x-2)^{m+1}}{(m+1)!}\right| \\ &= \frac{m!}{|\tilde{x}|^{m+1}}\cdot\frac{|x-2|^{m+1}}{(m+1)!} \\ &\leq \frac{1}{(m+1)|\tilde{x}|^{m+1}}. \end{aligned}$$

Now note that either $1<x\leq\tilde{x}\leq 2$ or $2\leq\tilde{x}\leq x<3$. In either case we have $|\tilde{x}|>1$, which implies that

$$\left|R_m(x)\right| \leq \frac{1}{(m+1)|\tilde{x}|^{m+1}} < \frac{1}{m+1},$$

and the right hand side converges to zero as $m\to\infty$. It follows that $\lim_{m\to\infty} R_m(x)=0$, so that

$$\ln(x)=\ln(2)+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n\cdot 2^n}(x-2)^n \quad \text{for } |x-2|<1.$$

To conclude this section, we provide a proof of Taylor's Theorem. The proof requires the following two results, which should be familiar from your first calculus course:

THEOREM 4.53 (Intermediate Value Theorem). *Suppose that the real function* $f\colon[a,b]\to\mathbb{R}$ *is continuous on the closed interval* $[a,b]$. *Moreover, suppose that* r *is a real number between* $f(a)$ *and* $f(b)$. *Then there exists a point* $\tilde{x}$ *in* $[a,b]$ *such that* $f(\tilde{x})=r$.

THEOREM 4.54 (Extreme Value Theorem). *Suppose that the real function* $f\colon[a,b]\to\mathbb{R}$ *is continuous on the closed interval* $[a,b]$. *Then* f *achieves a maximum and a minimum on* $[a,b]$. *That is, there exist points* t_1 *and* t_2 *in* $[a,b]$ *such that for all* t *in* $[a,b]$ *we have*

$$\begin{aligned} f(t) &\geq f(t_1) = \text{minimum of } f \text{ on } [a,b] \\ f(t) &\leq f(t_2) = \text{maximum of } f \text{ on } [a,b]. \end{aligned}$$

Proof of Taylor's Theorem. Assume that $x > c$ (the proof for $x < c$ is similar, but care must be taken about signs). By the Fundamental Theorem of Calculus, we have

$$f(x) = f(c) + \int_c^x f'(t)dt.$$

Now use integration by parts with $u = f'(t)$ and $v = t - x$ (so $dv = dt$):

$$\begin{aligned} f(x) &= f(c) + uv|_c^x - \int_c^x vdu \\ &= f(c) - f'(c)(c-x) - \int_c^x (t-x)f''(t)dt \\ &= f(c) + f'(c)(x-c) + \int_c^x f''(t)(x-t)dt. \end{aligned}$$

Repeat the integration by parts on the remaining integral, this time with $u = f''(t)$ and $dv = (x-t)dt$:

$$\begin{aligned} \int_c^x f''(t)(x-t)dt &= uv|_c^x - \int_c^x vdu \\ &= \frac{f''(c)}{2}(x-c)^2 + \int_c^x f'''(t)\frac{(x-t)^2}{2}dt. \end{aligned}$$

Putting this together with the previous computation, we find that

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2 + \int_c^x f'''(t)\frac{(x-t)^2}{2}dt.$$

Continuing to use integration by parts in the fashion, we find that

$$f(x) = T_m(x) + \int_c^x f^{(m+1)}(t)\frac{(x-t)^m}{m!}dt.$$

Subtracting the mth Taylor polynomial from both sides, we find an integral expression for the mth Taylor remainder of f at c:

$$R_m(x) = \int_c^x f^{(m+1)}(t)\frac{(x-t)^m}{m!}dt.$$

Now note that, by assumption, $f^{(m+1)}$ is a continuous function, hence it attains its minimum and maximum on the closed interval $[c, x]$ by the Extreme Value Theorem. Choose t_1 and t_2 in $[c, x]$ such that

$$f^{(m+1)}(t_1) = \text{minimum on } [c, x]$$

and

$$f^{(m+1)}(t_2) = \text{maximum on } [c, x].$$

It follows that

$$f^{(m+1)}(t_1) \int_c^x \frac{(x-t)^m}{m!} dt \leq R_m(x) \leq f^{(m+1)}(t_2) \int_c^x \frac{(x-t)^m}{m!} dt.$$

Evaluating the integrals yields

$$f^{(m+1)}(t_1) \frac{(x-c)^{m+1}}{(m+1)!} \leq R_m(x) \leq f^{(m+1)}(t_2) \frac{(x-c)^{m+1}}{(m+1)!}.$$

Since $f^{(m+1)}(t) \frac{(x-c)^{m+1}}{(m+1)!}$ is a continuous function of t, the Intermediate Value Theorem implies that there is a point $\tilde{x}$ between t_1 and t_2 such that

$$R_m(x) = f^{(m+1)}(\tilde{x}) \frac{(x-c)^{m+1}}{(m+1)!}.$$

□

Key points from Section 4.8:

- Power series centered at c (Definition 4.45)
- Taylor series, polynomials, and remainders (Definition 4.47)
- Taylor's Theorem (Theorem 4.49)
- Using Taylor's Theorem to show that $f(x)$ is represented by its Taylor series (Examples 4.50, 4.51)

4.9. Optional: Applications of Power Series

This section presents five applications of power series:

(1) The number e is irrational (Theorem 4.55)
(2) Numerical approximation of values of the natural logarithm and other functions (Example 4.57 and Exercise 4.12)
(3) The computation of non-elementary antiderivatives (Example 4.58)
(4) Limit computations for indeterminate forms (Example 4.59)
(5) The General Binomial Theorem (Example 4.60)

In Example 4.50, we showed that the real exponential function e^x is represented by its Maclaurin series:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

If we plug in $x = 1$, we find a series expression for the number e:

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

(In fact, this is really the *definition* of e, although you probably first encountered this number in a different way.)

THEOREM 4.55. *The real number $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ is irrational.*

PROOF. We begin by studying the error in the partial sum approximations s_m:

$$\begin{aligned}
\text{error}(m) &= e - s_m \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} - \sum_{n=0}^{m} \frac{1}{n!} \\
&= \sum_{n=m+1}^{\infty} \frac{1}{n!} \\
&= \frac{1}{(m+1)!} + \frac{1}{(m+2)!} + \frac{1}{(m+3)!} + \cdots \\
&= \frac{1}{(m+1)!} \left(1 + \frac{1}{m+2} + \frac{1}{(m+2)(m+3)} + \cdots \right).
\end{aligned}$$

Let $\sum_{k=0}^{\infty} a_k$ denote the series in parentheses on the last line, with $a_0 = 1$ and for $k \geq 1$

$$a_k = \frac{1}{(m+2)(m+3)\cdots(m+k+1)}.$$

Note that we have $a_k \leq 1/(m+1)^k$ for all $k \geq 0$, which means that the series $\sum_{k=0}^{\infty} a_k$ is bounded by the convergent geometric series

$$\sum_{k=0}^{\infty} \frac{1}{(m+1)^k} = \frac{1}{1 - \frac{1}{m+1}} = \frac{m+1}{m}.$$

Returning to the error term, we have

$$\text{error(m)} = \frac{1}{(m+1)!}\sum_{k=0}^{\infty} a_k \leq \frac{m+1}{m\cdot(m+1)!} = \frac{1}{m\cdot m!}.$$

Now suppose, in order to get a contradiction, that $e = \frac{p}{q}$ is a rational number, with p, q positive integers. Moreover, replacing p and q by $2p$ and $2q$ if necessary, we may assume that $q \geq 2$. Consider the partial sum s_q, with index given by the denominator of p/q. Our computation tells us that

$$e - s_q = \text{error}(q) \leq \frac{1}{q\cdot q!}$$

Multiplying both sides by the integer $q!$ we find that

$$0 < q!(e - s_q) \leq \frac{1}{q} < 1.$$

But note that $q!s_q$ is a positive integer:

$$\begin{aligned} q!s_q &= q!\left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{(q-1)!} + \frac{1}{q!}\right) \\ &= q! + q! + q(q-1)\cdots 3 + q(q-1)\cdots 4 + \cdots + q + 1. \end{aligned}$$

This implies that $q!(e - s_q)$ is also a positive integer:

$$q!(e - s_q) = q!\frac{p}{q} - q!s_q = p\cdot(q-1)! - q!s_q.$$

But this is a contradiction, since there are no integers between 0 and 1.

□

Remark 4.56. In the proof of the previous theorem, we established a bound for the error in the partial sum approximations of e:

$$\text{error}(m) = e - s_m \leq \frac{1}{m\cdot m!}.$$

The numbers on the right hand side approach zero rapidly, so the partial sum approximations s_m are quite good even for small values of m. For instance, taking $m = 9$ we find that

$$\frac{1}{9\cdot 9!} \approx 3.06\times 10^{-7} = 0.000000306.$$

It follows that the partial sum s_9 approximates e correctly to 5 decimal places:

$$e \approx \sum_{n=0}^{9} \frac{1}{n!} \approx 2.71828$$

As the previous remark indicates, another application of power series to to provide approximations to numbers of interest. In particular, when we have an alternating series with decreasing term-sizes, then we can use the technique described in Example 3.38 to estimate the sum, with good control over the error.

EXAMPLE 4.57. Suppose that we want to find an approximation of $\ln(1.1)$, correct to 4 decimal places. Consider the Maclaurin series representation for $\ln(1+x)$:

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{(n+1)} \frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

This formula is valid for $|x| < 1$, and in particular at $x = 0.1$:

$$\ln(1.1) = (0.1) - \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} - \frac{(0.1)^4}{4} + \cdots$$

Because this series is alternating with decreasing term-sizes, the error in the mth partial sum approximation is bounded by the size of the $(m+1)$st term. In particular, if we keep the first four terms displayed above, then the error is at most $(0.1)^5/5 = 0.000002$. Hence, the 4th partial sum approximation is correct to 4 decimal places:

$$\ln(1.1) = (0.1) - \frac{(0.1)^2}{2} + \frac{(0.1)^3}{3} - \frac{(0.1)^4}{4} \approx 0.0953$$

EXERCISE 4.12. Compute a decimal approximation to $\arctan(0.5)$ correct to two decimal places, using the Maclaurin series formula

$$\arctan(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots .$$

Our next application concerns the computation of antiderivatives.

EXAMPLE 4.58. Consider the real function $f(x) = e^x/x$. You may be surprised to learn that this innocent looking function has no *elementary* antiderivative. That is, there is no finite combination (using addition, subtraction, multiplication, division, and composition) of familiar functions (such as polynomials, nth roots, exponentials, logarithms, trigonometric or inverse trigonometric functions) with derivative equal to $f(x)$. Nevertheless, since $f(x)$ is a continuous function, the Fundamental Theorem of Calculus guarantees that it *does* possess an antiderivative:

$$F(x) = \int_1^x \frac{e^t}{t} dt.$$

This function satisfies $F(1) = 0$ and $F'(x) = f(x) = e^x/x$ for all $x > 0$. As a theoretical description of the antiderivative, the integral is fine. But for further analysis and the computation of specific values, it would be better to have a more explicit formula, and this is where power series come in.

Note that we can obtain a series formula for $f(x) = e^x/x$ by dividing the Maclaurin series for e^x by x:

$$\begin{aligned} \frac{e^x}{x} &= \frac{1}{x}\sum_{n=0}^{\infty} \frac{x^n}{n!} \\ &= \frac{1}{x}\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right) \\ &= \frac{1}{x} + 1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots \\ &= \frac{1}{x} + \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}. \end{aligned}$$

Thus, we have represented $f(x)$ as the reciprocal function $1/x$ plus a power series, and the formula is valid for all $x > 0$. But now we can compute the antiderivative $F(x)$ term-by-term:

$$\begin{aligned} F(x) &= C + \ln(x) + \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)\cdot(n+1)!} \\ &= C + \ln(x) + x + \frac{x^2}{2\cdot 2!} + \frac{x^3}{3\cdot 3!} + \cdots. \end{aligned}$$

Any choice for the constant of integration C yields an antiderivative for e^x/x; if we want to obtain the particular antiderivative $F(x)$ satisfying $F(1) = 0$, we must choose $C = -\sum_{n=1}^{\infty} \frac{1}{n \cdot n!}$.

Now we provide an application of power series to the computation of function limits.

EXAMPLE 4.59. Consider the following limit:

$$\lim_{x \to 0} \left(\frac{\sin(x^2)}{x^4} - \frac{\cos(x)}{x^2} \right).$$

In your first calculus course you might have used L'Hôpital's Rule multiple times for this problem. But we will use power series instead. We begin by combining the fractions and then finding a power series formula for the numerator, starting with the Maclaurin series for $\sin(x)$ and $\cos(x)$.

$$\frac{\sin(x^2)}{x^4} - \frac{\cos(x)}{x^2} = \frac{\sin(x^2) - x^2\cos(x)}{x^4}.$$

To understand the numerator as a power series, write

$$\begin{aligned} \sin(x^2) &= \sum_{n=0}^{\infty} (-1)^n \frac{(x^2)^{2n+1}}{(2n+1)!} = x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \cdots \\ x^2\cos(x) &= x^2 \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = x^2 - \frac{x^4}{2!} + \frac{x^6}{4!} - \frac{x^8}{8!} + \cdots \end{aligned}$$

Now subtract, and note that the degree-2 terms cancel:

$$\sin(x^2) - x^2\cos(x) = \frac{x^4}{2} - \left(\frac{1}{6} + \frac{1}{24} \right) x^6 - \frac{x^8}{8!} + \cdots.$$

Dividing by x^4, we get a power series formula for our original expression:

$$\frac{\sin(x^2) - x^2\cos(x)}{x^4} = \frac{1}{2} - \frac{5}{24}x^2 - \frac{x^4}{8!} + \cdots.$$

Finally, we take the limit, and observe that all terms after the first on the right hand side go to zero:

$$\begin{aligned}\lim_{x\to 0}\left(\frac{\sin(x^2)-x^2\cos(x)}{x^4}\right) &= \lim_{x\to 0}\left(\frac{1}{2}-\frac{5}{24}x^2-\frac{x^4}{8!}+\cdots\right)\\ &= \frac{1}{2}.\end{aligned}$$

Note that, in the final step, we have used the fact that the power series is continuous at $x=0$; this follows from the fact that it is differentiable.

EXAMPLE 4.60 (General Binomial Theorem). Fix a real exponent p, and consider the real function $f(x)=(1+x)^p$. This function is defined and infinitely differentiable near $x=0$, with

$$\begin{aligned} f(x) &= (1+x)^p\\ f'(x) &= p(1+x)^{p-1}\\ f''(x) &= p(p-1)(1+x)^{p-2}\\ &\vdots\\ f^{(n)}(x) &= p(p-1)\cdots(p-n+1)(1+x)^{p-n}\\ &\vdots \end{aligned}$$

Evaluating at $x=0$ and dividing by $n!$ yields the coefficients of the Maclaurin series:

$$a_0=1,\ a_1=p,\ a_2=\frac{p(p-1)}{2},\ldots,a_n=\frac{p(p-1)\cdots(p-n+1)}{n!},\ldots$$

Note that, in the special case $p=m$ an integer, we have

$$a_n=\frac{m!}{n!(m-n)!}=\binom{m}{n},$$

the binomial coefficient. For this reason, for arbitrary real numbers p we *define* the symbol

$$\binom{p}{n}=\frac{p(p-1)\cdots(p-n+1)}{n!}.$$

Using this notation, we find that the Maclaurin series for $f(x)$ is given by

$$\sum_{n=0}^{\infty} \binom{p}{n} x^n.$$

First note that if $p = m \geq 0$ is a nonnegative integer, then the series is actually a finite polynomial, equal to $(1+x)^m$ by the ordinary binomial theorem. So assume that m is not a nonnegative integer, which implies that all coefficients a_n are nonzero. To find the radius of convergence R, we use the ratio test:

$$\begin{aligned}
\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} &= \lim_{n\to\infty} \frac{|p(p-1)\cdots(p-n)|}{|p(p-1)\cdots(p-n+1)|} \cdot \frac{n!}{(n+1)!} \\
&= \lim_{n\to\infty} \frac{|p-n|}{n+1} \\
&= \lim_{n\to\infty} \frac{|1-\frac{p}{n}|}{1+\frac{1}{n}} \\
&= 1 \\
&= L,
\end{aligned}$$

so $R = 1/L = 1$, and the series converges absolutely for $|x| < 1$. We wish to show that in fact, the series converges to the original function $f(x) = (1+x)^p$ for $|x| < 1$.

At this point, all we know is that the Maclaurin series converges to *some* function $s(x)$ that is infinitely differentiable on the interval $-1 < x < 1$. Moreover, the derivative of this function is

$$s'(x) = \sum_{n=0}^{\infty} n \binom{p}{n} x^{n-1},$$

so if we multiply by $1+x$ we find that

$$\begin{aligned}
(1+x)s'(x) &= s'(x) + xs'(x) \\
&= \sum_{n=0}^{\infty} n\binom{p}{n}x^{n-1} + x\sum_{n=0}^{\infty} n\binom{p}{n}x^{n-1} \\
&= \sum_{n=0}^{\infty} n\binom{p}{n}x^{n-1} + \sum_{n=0}^{\infty} n\binom{p}{n}x^{n} \\
&= \sum_{n=0}^{\infty}(n+1)\binom{p}{n+1}x^{n} + \sum_{n=0}^{\infty} n\binom{p}{n}x^{n} \\
&= \sum_{n=0}^{\infty}\left((n+1)\binom{p}{n+1} + n\binom{p}{n}\right)x^{n}.
\end{aligned}$$

Let's investigate that complicated looking coefficient:

$$\begin{aligned}
(n+1)\binom{p}{n+1} + n\binom{p}{n} &= (n+1)\frac{p(p-1)\cdots(p-n)}{(n+1)!} + n\binom{p}{n} \\
&= \frac{p(p-1)\cdots(p-n)}{n!} + n\binom{p}{n} \\
&= (p-n)\frac{p(p-1)\cdots(p-n+1)}{n!} + n\binom{p}{n} \\
&= (p-n)\binom{p}{n} + n\binom{p}{n} \\
&= p\binom{p}{n}.
\end{aligned}$$

Returning to the previous computation, we find that

$$(1+x)s'(x) = \sum_{n=0}^{\infty} p\binom{p}{n}x^{n} = ps(x).$$

Now consider the function $g(x) = (1+x)^{-p}s(x)$, which (being the product of two differentiable functions) is differentiable on the interval $-1 < x < 1$. We compute the derivative using first the product rule and then the previous computation:

$$\begin{aligned} g'(x) &= -p(1+x)^{-p-1}s(x) + (1+x)^{-p}s'(x) \\ &= -(1+x)^{-p-1}(1+x)s'(x) + (1+x)^{-p}s'(x) \\ &= -(1+x)^{-p}s'(x) + (1+x)^{-p}s'(x) \\ &= 0. \end{aligned}$$

Since the derivative of g is zero for all x, it follows that $g(x) = C$ is a constant function, so that $s(x) = C(1+x)^p$. Plugging in $x = 0$ reveals that the constant $C = 1$, so that $s(x) = (1+x)^p$ as expected:

$$(1+x)^p = \sum_{n=0}^{\infty} \binom{p}{n} x^n.$$

This is the *General Binomial Theorem*.

We now consider the complex function $F(z)$ defined by the same Maclaurin series:

$$F(z) = \sum_{n=0}^{\infty} \binom{p}{n} z^n.$$

The function $F(z)$ is absolutely convergent on the open unit disc, where it defines a complex differentiable function that extends the real function $(1+x)^p$. Moreover, $F(z)$ is the *only* possible complex differentiable extension of $(1+x)^p$, and for this reason we denote it by $(1+z)^p$. In this way, we have extended the use of arbitrary real exponents p to the complex plane, at least for points $1+z$ in the open disc of radius 1 centered at 1.

EXERCISE 4.13. Consider the case $p = -(m+1) < 0$ a negative exponent. Replace z by $-z$ in the general binomial theorem, and check that the series formula you obtain for the function $1/(1-z)^{m+1}$ matches the one we obtained from the geometric series in Example 4.27.

Key points from Section 4.9:

- The number e is irrational (Thm 4.55)
- Series approximation of function values (Example 4.57)
- Power series for antiderivatives (Example 4.58)
- Using power series to compute function limits (Example 4.59)
- General Binomial Theorem (Example 4.60)

4.10. Optional: Back to the Riemann Zeta Function

We have some unfinished business from way back in Section 3.1. There, we began our discussion of series by writing down the Riemann zeta function:

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z} = 1 + \frac{1}{2^z} + \frac{1}{3^z} + \ldots,$$

but we immediately retreated to consideration of real inputs $z = p$:

$$\zeta(p) = \sum_{n=1}^{\infty} \frac{1}{n^p} = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \ldots.$$

We used the integral test in Section 3.3 to show that the p-series defining $\zeta(p)$ converges for $p > 1$ and diverges to $+\infty$ for $p \leq 1$. We are now ready to consider general complex inputs z to the zeta function.

For this, we need to make sense of the expression n^z for n a positive integer and z a complex number. This is an example of our complex extension problem: we want to extend the real function $f(p) = n^p$ to a complex function $F(z)$. We begin by looking carefully at the real case, using the natural logarithm to rewrite the function $f(p) = n^p$ in terms of the real exponential:

$$f(p) = n^p = e^{p \ln(n)} = \exp(p \ln(n)).$$

We can now replace the real variable p with the complex variable z, obtaining a complex function $F(z)$:

$$F(z) = \exp(z \ln(n)).$$

This function is defined for all complex z and satisfies $F(p) = n^p$ for all real numbers p. That is, $F(z)$ is is the unique complex differentiable extension of the real function $f(p) = n^p$ to the complex plane. For this reason, we write $n^z = F(z) = \exp(z\ln(n))$. Note that since $\exp(w)\exp(-w) = \exp(w-w) = \exp(0) = 1$, we have

$$\frac{1}{n^z} = \frac{1}{\exp(z\ln(n))} = \exp(-z\ln(n)).$$

With this understanding of the meaning of the terms $1/n^z$, we can finally understand the definition of the Riemann zeta function:

$$\begin{aligned}\zeta(z) &= \sum_{n=1}^{\infty}\frac{1}{n^z}\\ &= \sum_{n=1}^{\infty}\exp(-z\ln(n))\\ &= 1+\exp(-z\ln(2))+\exp(-z\ln(3))+\cdots.\end{aligned}$$

To determine where this series converges, set $z = x+iy$, and note that the nth term is

$$\exp(-z\ln(n)) = \exp(-(x+iy)\ln(n)) = \exp(-x\ln(n))\exp(-iy\ln(n)).$$

Moreover, because $\exp(-iy\ln(n))$ lies on the unit circle, the magnitude of the nth term is

$$\big|\exp(-z\ln(n))\big| = e^{-x\ln(n)} = \frac{1}{n^x}.$$

So the series of magnitudes is given by

$$\sum_{n=1}^{\infty}\frac{1}{|n^z|} = \sum_{n=1}^{\infty}\frac{1}{n^x}$$

which converges for $x > 1$. Thus, the series defining the Riemann zeta function converges absolutely for all complex z with real part greater than 1. Figure 4.11 shows the graph of the magnitude $|\zeta(z)|$ for $z = x+iy$ with $x > 1$. The large yellow spike in the center represents the divergence of the harmonic series.

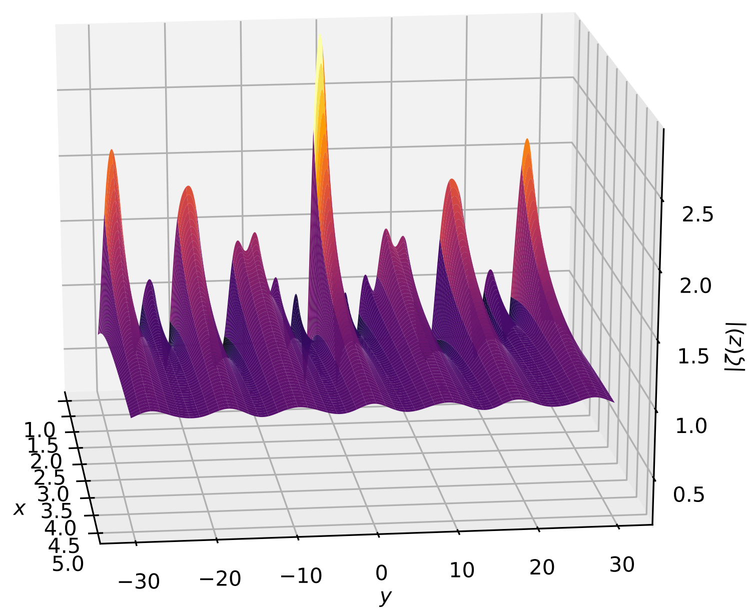


FIGURE 4.11. Graph of the magnitude of the Riemann zeta function.

Note that the series formula for $\zeta(z)$ is not a *power* series formula: the terms are complex exponentials $\exp(-z\ln(n))$ rather than monomials $a_n z^n$, and the series does not look like an "infinite polynomial." So our results about power series do not tell us whether $\zeta(z)$ is differentiable, for instance. Although we are not in a position to prove it here, $\zeta(z)$ *is* complex differentiable, and provides one of the most important and mysterious functions in all of mathematics.

The domain of the Riemann zeta function $\zeta(z)$ may be extended to include the entire complex plane $\mathbb{C}$ with the exception of the single point $z = 1$ (where it diverges as the harmonic series). Actually, we saw an elementary instance of this "expansion of the domain" in our very first example of the geometric series. Namely, even though the geometric series $\sum_{n=0}^{\infty} z^n$ converges only on the open disc $|z| < 1$, it defines the function $1/(1-z)$ on that disc, and that formula makes sense for all $z \neq 1$. In a similar but more sophisticated way, $\zeta(z)$

extends to a function that makes sense for all $z \neq 1$, even though it is no longer defined by the original series on the larger domain. This extension method is called *analytic continuation*, and you will learn about it if you take MATH 535: Complex Analysis.

The famous Riemann Hypothesis mentioned in Section 3.1 concerns the locations of the zeros of the extended zeta function $\zeta(z)$. It turns out that the zeta function vanishes at all negative even integers:

$$\zeta(2n) = 0 \qquad \text{for all integers } n < 0.$$

These zeros are well-understood, and for that reason are called the *trivial* zeros of $\zeta(z)$. But the other zeros are mysterious: it is known that all non-trivial zeros lie in the vertical *critical strip* defined by $0 < \mathrm{Re}(z) < 1$. The Riemann Hypothesis is the statement that they actually lie on the *critical line* defined by $\mathrm{Re}(z) = 1/2$:

Riemann Hypothesis: If w is a nontrivial zero of the Riemann zeta function, then $\mathrm{Re}(w) = \frac{1}{2}$.

The import of this statement for number theory is highly non-obvious, but may be roughly stated as follows: the prime numbers p are distributed in the "best possible way" among all the integers.

The example of the Riemann zeta function illustrates that there are other important types of series besides power series. In particular, $\zeta(z)$ is the most important instance of a *Dirichlet series*, which have the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^z} = a_1 + \frac{a_2}{2^z} + \frac{a_3}{3^z} + \cdots ,$$

for complex coefficients a_n. These series play a central role in the subject of analytic number theory.

Trigonometric series are real series of the form

$$\sum_{n=0}^{\infty} (a_n \cos(n\theta) + b_n \sin(n\theta)).$$

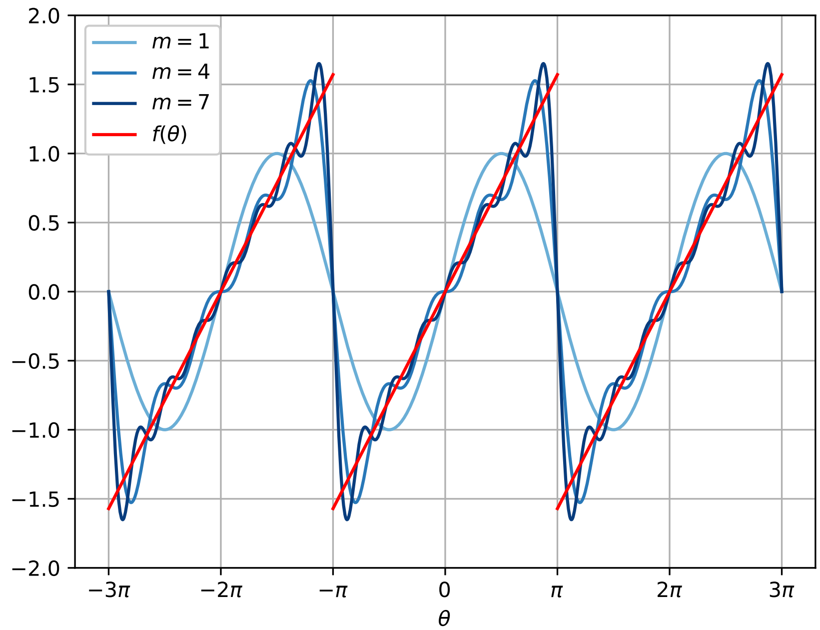


FIGURE 4.12. Partial sums of the trigonometric series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(n\theta)$ together with the periodic extension $f(\theta)$ of the linear function $\theta/2$ on the interval $(-\pi, \pi)$.

Here, θ is a real variable and the coefficients a_n and b_n are real. Note that each term has period 2π, so these series define periodic functions $f\colon D \to \mathbb{R}$, where D is the domain of convergence. Trigonometric series are also known as *real Fourier series*. Figure 4.12 shows some partial sums of the trigonometric series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(n\theta) = \sin(\theta) - \frac{1}{2}\sin(2\theta) + \frac{1}{3}\sin(3\theta) - \cdots,$$

which seem to be converging to the function $f(\theta)$ obtained by periodic extension of the linear function $\theta/2$ on $(-\pi, \pi)$.

There are also *complex Fourier series*, which have the following form (where we have written $e^{in\theta} = \exp(in\theta)$):

$$\sum_{n=-\infty}^{\infty} c_n e^{in\theta} = c_0 + c_1 e^{i\theta} + c_{-1} e^{-i\theta} + c_2 e^{2i\theta} + c_{-2} e^{-2i\theta} + \cdots.$$

Here, θ is still a real variable, but the coefficients c_n are complex, so these series define complex-valued periodic functions $g\colon D \to \mathbb{C}$, where $D \subseteq \mathbb{R}$ is the real domain of convergence.

The subject of *Fourier analysis* addresses questions similar to those we asked about power series in this chapter:

(1) what are the domains of convergence D?
(2) are the functions defined by Fourier series differentiable?
(3) given a periodic function $G(\theta)$, can we find a Fourier series that represents it?

The answers to these questions are all more subtle than in the case of power series, and require a deeper study of the concept of convergence. But the answers are well worth the effort to discover: just like power series, Fourier series have important applications to many areas of pure and applied mathematics, physics, and engineering.

Key points from Section 4.10:

- Meaning of the complex function n^z in terms of the complex exponential function (page 239)

4.11. In-text Exercises

This section collects the in-text exercises that you should have worked on while reading the chapter.

EXERCISE 4.1 Spend some time pondering the picture on page 174. Can you convince yourself that f takes vertical lines to circles, as the picture indicates? To get started, consider the imaginary axis $z = iy$. Then

$$f(iy) = \frac{1}{1-iy} = \frac{1+iy}{1+y^2}.$$

As y varies, do you see why $f(iy)$ traces out the black circle on the right hand side? Investigate the other vertical lines in a similar manner.

EXERCISE 4.2 For $f(x) = 1/(1-x)$ and the third partial sum $s_3(x) = 1 + x + x^2 + x^3$, verify by explicit computation that

$$f(0) = s_3(0), \quad f'(0) = s_3'(0), \quad f''(0) = s_3''(0), \quad f^{(3)}(0) = s_3^{(3)}(0).$$

EXERCISE 4.3 Adapt the argument given in Example 4.9 to show that the power series below has radius of convergence $R = +\infty$.

$$\sum_{n=0}^{\infty}(-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots.$$

EXERCISE 4.4 Fill in the details in the proof of Proposition 4.17.

EXERCISE 4.5 Check that the power series for $f'(z)$ in Example 4.20 also has radius of convergence $R = 1/2$.

EXERCISE 4.6 Verify that the antiderivative power series $F(z)$ in Example 4.23 has radius of convergence $R = 1/2$.

EXERCISE 4.7 Let $L(z)$ and $A(z)$ denote the power series from Examples 4.28 and 4.32:

$$L(z) = -\sum_{n=0}^{\infty} \frac{z^{n+1}}{n+1} = -z - \frac{z^2}{2} - \frac{z^3}{3} - \cdots$$

and

$$A(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{2n+1} = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \cdots$$

Show that for $|z| < 1$, we have

$$A(z) = \frac{i}{2}\left(L(iz) - L(-iz)\right).$$

EXERCISE 4.13 Consider the case $p = m + 1 < 0$ a negative exponent. Replace z by $-z$ in the general binomial theorem, and check that the series formula you obtain for the function $1/(1-z)^{m+1}$ matches the one we obtained from the geometric series in Example 4.27.

EXERCISE 4.8 Compute the Maclaurin series for $\cos(x)$, and show that the complex version matches the series from Example 4.9, with radius of convergence $R = \infty$:

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

EXERCISE 4.9 (Euler's Formula) Recall the power series defining the complex exponential, sine, and cosine as functions on the complex plane:

$$\begin{aligned}
\exp(z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \\
\sin(z) &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots \\
\cos(z) &= \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots
\end{aligned}$$

Using these power series, show that for all complex numbers z,

$$\exp(iz) = \cos(z) + i\sin(z).$$

Exercise 4.10 Use de Moivre's formula to find identities for $\cos(5\theta)$ and $\sin(5\theta)$ in terms of $\cos(\theta)$ and $\sin(\theta)$.

Exercise 4.11 Use Taylor's Theorem to show that $\cos(x)$ is represented by its Maclaurin series for all real x.

Exercise 4.12 Compute a decimal approximation to $\arctan(0.5)$ correct to two decimal places, using the Maclaurin series formula

$$\arctan(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{2n+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots .$$

Exercise 4.13 Consider the case $p = -(m+1) < 0$ a negative exponent. Replace z by $-z$ in the general binomial theorem, and check that the series formula you obtain for the function $1/(1-z)^{m+1}$ matches the one we obtained from the geometric series in Example 4.27.

4.12. Problems

4.1. For each of the following, check that the polynomial $p(x)$ is the best mth degree approximation of $f(x)$ near $x = c$, for the specified values of m and c (see EXERCISE 4.2):

(a) $f(x) = \arctan(2x)$, $p(x) = 2x - \frac{8}{3}x^3$; $m = 3, c = 0$

(b) $f(x) = \sin(x)+\cos(x)$, $p(x) = 1+x-\frac{1}{2}x^2-\frac{1}{6}x^3+\frac{1}{24}x^4$; $m = 4, c = 0$

(c) $f(x) = x^{2/3}$, $p(x) = 1+\frac{2}{3}(x-1)-\frac{1}{9}(x-1)^2+\frac{4}{81}(x-1)^3$; $m = 3, c = 1$

4.2. For each of the following coefficient sequences $(a_n)_{n\geq 0}$, write out the corresponding power series using both the summation notation and the expanded $+\cdots$ form including at least 5 terms:

$$\sum_{n=0}^{\infty} a_n z^n = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + \cdots$$

(a) $(a_n) = (1)$

(b) $(a_n) = \left(\frac{\sqrt{2n+1}}{n!}\right)$

(c) $(a_n) = \left(\frac{(-1)^n \pi^{2n}}{\ln(n+2)}\right)$

4.3. Find the radius of convergence for each of the following power series:

(a) $\sum_{n=0}^{\infty} \sqrt{n} z^n$

(b) $\sum_{n=1}^{\infty} \frac{2^n z^n}{n^3}$

(c) $\sum_{n=1}^{\infty} n^n z^n$

(d) $\sum_{n=1}^{\infty} \frac{z^n}{n^n}$

(e) $\sum_{n=1}^{\infty} \frac{n^3 z^n}{3^n}$

(f) $\sum_{n=2}^{\infty} (-1)^n \frac{z^n}{3^n \ln(n)}$

(g) $\sum_{n=1}^{\infty} 5^n \sqrt{n} z^n$

(h) $\sum_{n=0}^{\infty} \frac{(3n)! z^n}{n!(2n)!}$

4.4. Find the radius of convergence of the following power series:

$$\sum_{n=0}^{\infty} \frac{(n!)^{17}}{(17n)!} z^n$$

4.5. Use the "missing terms method" illustrated in Example 4.9 to find the radius of convergence of the following power series:

$$\sum_{n=2}^{\infty} \frac{z^{2n}}{2^n (\ln n)^2}$$

4.6. Let (a_n) be a coefficient sequence and suppose that the limit $A = \lim_{n\to\infty} \frac{|a_n|}{|a_{n+1}|}$ exists. Determine the radius of convergence of each of the following power series:

(a) $\sum_{n=0}^{\infty} a_n z^n$
(b) $\sum_{n=0}^{\infty} a_n z^{2n}$
(c) $\sum_{n=0}^{\infty} a_n^2 z^n$

4.7. Consider the power series:

$$\sum_{n=0}^{\infty} \frac{z^{(2^n)}}{2^n} = z + \frac{z^2}{2} + \frac{z^4}{4} + \frac{z^8}{8} + \cdots$$

(a) Show that the series converges absolutely for $|z| \le 1$.
(b) Show that if $|z| > 1$, then the series diverges. [Hint: use the divergence test.]

4.8. For each of the following power series, first find the radius of convergence, and then compute the derivative as a power series. What is the radius of convergence of the derivative?

(a) $\sum_{n=0}^{\infty} nz^n$
(b) $\sum_{n=1}^{\infty} \frac{z^n}{\sqrt{n}}$
(c) $\sum_{n=1}^{\infty} \frac{z^n}{(n+1)^n}$
(d) $\sum_{n=1}^{\infty} \frac{z^n}{n^3}$
(e) $\sum_{n=1}^{\infty} n^3 z^n$
(f) $\sum_{n=0}^{\infty}(-1)^n \frac{z^n}{2^n}$
(g) $\sum_{n=1}^{\infty} 5^n z^n$

4.9. For each of the power series in Problem 4.8, compute the antiderivative as a power series. What is the radius of convergence of the antiderivative?

4.10. For each of the following functions, find a power series representation by using a geometric series expansion. Determine the radius of convergence.

(a) $\frac{1}{1+z}$
(b) $\frac{2}{1-z^3}$
(c) $\frac{1}{z+5}$
(d) $\frac{1+z}{1-z}$
(e) $\frac{z^2}{1-z^4}$
(f) $\frac{2z}{3z^2+2}$

4.11. Consider the function $f(z) = \frac{3}{z^2+z-2}$.

(a) What is the domain of f?

(b) Find constants A and B such that $f(z) = \frac{A}{z-1} + \frac{B}{z+2}$.

(c) Use part (b) to find a power series representation of $f(z)$.

(d) What is the radius of convergence of the power series you found in part (c)?

4.12. For each of the following real functions, find a power series representation in two different ways: (i) by immediately using a geometric series expansion; (ii) by recognizing the function as the derivative of a logarithm—Example 4.28 may be helpful. Determine the radius of convergence.

(a) $\frac{-2x}{1-x^2}$

(b) $\frac{x}{4+x^2}$

4.13. Fix a positive integer $k \geq 2$. In this problem, you will find an exact value for the infinite series $\sum_{n=1}^{\infty} \frac{n}{k^n}$ and $\sum_{n=2}^{\infty} \frac{n^2}{k^n}$. Recall that Example 4.27 shows that for $|z| < 1$ we have

$$\sum_{n=1}^{\infty} nz^{n-1} = \frac{1}{(1-z)^2} \qquad \text{and} \qquad \sum_{n=1}^{\infty} n(n-1)z^{n-2} = \frac{2}{(1-z)^3}.$$

(a) Evaluate at $z = \frac{1}{k}$ to find exact values for the series $\sum_{n=1}^{\infty} \frac{n}{k^n}$ and $\sum_{n=1}^{\infty} \frac{n(n-1)}{k^n}$.

(b) Now find the exact value for the series $\sum_{n=1}^{\infty} \frac{n^2}{k^n}$.

4.14. Use the method illustrated in Example 4.36 to compute the Maclaurin series of the function $f(x) = \sqrt[3]{2+x}$. Determine the radius of convergence.

4.15. Use the method illustrated in Example 4.36 to compute the Maclaurin series of the function $f(x) = \ln(1+x)$. Compare to the result of Example 4.28.

4.16. Use the method illustrated in Example 4.36 to compute the Maclaurin series of the function $f(x) = \ln(2+3x)$. Determine the radius of convergence.

4.17. Compute the Maclaurin series of the function $f(x) = \sin(\pi x)$ in two different ways:

(a) Using the method illustrated in Example 4.36.
(b) Using the result of Example 4.41 and the fact that $\sin(x)$ is equal to its Maclaurin series for all x.

4.18. Use the power series for e^x to compute the Maclaurin series of $\sqrt{e^x}$.

4.19. Use the power series for the exponential function to compute the Maclaurin series of the function $f(z) = z^3 \exp(2z^2)$.

4.20. This problem concerns the complex sine function $\sin(z)$, defined by the power series in Example 4.41.

(a) Find the Maclaurin series for the function $\frac{\sin(z)}{z}$.
(b) Use your answer to part (a) to compute the limit $\lim_{x\to 0} \frac{\sin(x)}{x}$.
(c) Use your answer to part (a) to compute the Maclaurin series for the function $\frac{\sin(z^2)}{z^2}$.
(d) Now find a power series antiderivative for the function $\frac{\sin(z^2)}{z^2}$.
(e) Finally, use your answer to part (d) to find an approximation to the following definite integral, correct to 3 decimal places:

$$\int_0^1 \frac{\sin(x^2)}{x^2} dx.$$

4.21. In this problem, you will derive the following identity for the sine and cosine functions, valid for all complex $z \neq 2\pi m$, an integer multiple of 2π:

$$\frac{1}{2} + \cos(z) + \cos(2z) + \cdots + \cos(nz) = \frac{\sin((n+\frac{1}{2})z)}{2\sin(\frac{z}{2})}.$$

(a) Show that for all complex w, we have $\cos(w) = \frac{1}{2}\big(\exp(iw) + \exp(-iw)\big)$.
(b) Use part (a) to show that

$$\frac{1}{2} + \cos(z) + \cos(2z) + \cdots + \cos(nz) = \frac{1}{2}\exp(-inz)\sum_{k=0}^{2n} \exp(ikz).$$

(c) Note that the sum on the right hand side in part (b) is the partial sum of a geometric series; use the formula $\sum_{k=0}^{2n} w^k = \frac{1-w^{2n+1}}{1-w}$ to rewrite the expression.

(d) Now use the formula $\sin(w) = \frac{1}{2i}\big(\exp(iw) - \exp(-iw)\big)$ to finish.

(e) Observe the following: for $z = x$ real, we have the following identity that makes no mention of complex numbers:

$$\frac{1}{2} + \cos(x) + \cos(2x) + \cdots + \cos(nx) = \frac{\sin((n+\frac{1}{2})x)}{2\sin(\frac{x}{2})}.$$

While this statement is entirely real, the *derivation* makes essential use of the complex trigonometric and exponential functions.

CHAPTER 5

APPLICATIONS TO DIFFERENTIAL EQUATIONS

5.1. Motivation: Population Growth

Suppose we have a population of bacteria in a petri dish, and we wish to study the number of cells in the population as time goes on. For this purpose, we introduce an unknown function $P(t)$ such that

$$P(t) = \text{ number of cells at time } t.$$

We imagine that the experiment starts at time $t = 0$, when there are some initial number P_0 of cells present. The population of bacteria will grow through the process of cell-division, and we assume that *each cell makes the same fixed contribution to the growth rate at each instant of time*. Our goal is to predict the number of cells $P(t)$ at later times $t > 0$.

Recall that the rate of change of a function is given by its derivative, so our assumption is really about the derivative $P'(t)$ of the unknown function $P(t)$. More precisely, our assumption says that the *relative* *growth rate* $P'(t)/P(t)$ is a positive constant $r > 0$:

$$\frac{P'(t)}{P(t)} = r \qquad \text{for all } t \geq 0.$$

In summary, we may express our assumptions about the unknown function $P(t)$ as follows:

$$P'(t) = rP(t) \quad \text{for all } t \geq 0, \qquad \text{and} \quad P(0) = P_0.$$

On the left we have a *differential equation*, and we seek a *solution*: a particular function $P(t)$ that makes the equation true. Moreover, we want to find a solution that also satisfies the *initial condition* $P(0) = P_0$. In this case, it is easy to write down a solution explicitly: $P(t) = Ce^{rt}$ where C is any constant. Indeed, by direct computation:

$$P'(t) = (Ce^{rt})' = Cre^{rt} = rP(t).$$

In order the match the initial condition $P(0) = P_0$, we must choose the constant $C = P_0$:

$$P(t) = P_0 e^{rt}.$$

In fact (Problem 5.1), the solution $P(t) = P_0 e^{rt}$ is the *only* solution to the differential equation satisfying the initial condition $P(0) = P_0$. We express this fact by saying that it is the *unique* solution with the specified initial condition. So, our assumption of a constant relative growth rate leads to the prediction that the population of bacteria will grow exponentially.

The differential equation $P'(t) = rP(t)$ shows up in many different contexts: in addition to population growth, it models certain chemical reactions as well as continuously compounded interest. And if the constant $r < 0$ is negative, then it describes population decline, radioactive decay, and other phenomena. This is an instance of a fact that we should marvel at: many seemingly unrelated physical and social systems may be successfully modeled by a single differential equation.

Of course, in our finite world, no growth can be exponential forever, and so we need more sophisticated assumptions to better represent the long-term growth of real populations. The following *logistic differential* *equation* models a population of individuals who must compete for resources:

$$P'(t) = rP(t) - \frac{r}{K}(P(t))^2.$$

Briefly, the first term $rP(t)$ on the right corresponds to our old assumption of constant relative growth, but the second term $-\frac{r}{K}(P(t))^2$ serves to suppress the growth rate as the population gets large. If there are $P(t)$ individuals at time t, then there are roughly $(P(t))^2$ opportunities for competitive interaction between individuals, and the assumption is that each possible interaction decreases the growth rate by the same constant amount $\frac{r}{K}$. The positive constant $K > 0$ is called the *carrying capacity*.

EXERCISE 5.1. Consider the logistic differential equation with relative rate $r = 0.1$ and carrying capacity $K = 100$:

$$P'(t) = 0.1P(t) - \frac{0.1}{100}(P(t))^2$$

a) By direct computation, verify that the following function is a solution satisfying the initial condition $P(0) = P_0$:

$$P(t) = \frac{P_0 e^{0.1t}}{1 + \frac{P_0}{100}(e^{0.1t} - 1)}$$

b) Show that if $P_0 > 0$, then $\lim_{t\to\infty} P(t) = 100$, so that in the long-run, the population described by this logistic equation approaches its carrying capacity $K = 100$.

Figure 5.1 shows several solutions to the logistic equation from the previous exercise (for different initial conditions P_0), and compares them to the simpler model of exponential growth.

In general, the type of story we told above about population growth is common: in applied problems we often have an unknown function $f(t)$ that represents some quantity of interest, and we have some reasonable assumptions about the relationship between $f(t)$ and its derivatives. The resulting differential equation then provides the starting point for the study of the quantity represented by $f(t)$. In the optional Section 5.3 we present an extended example of this method for

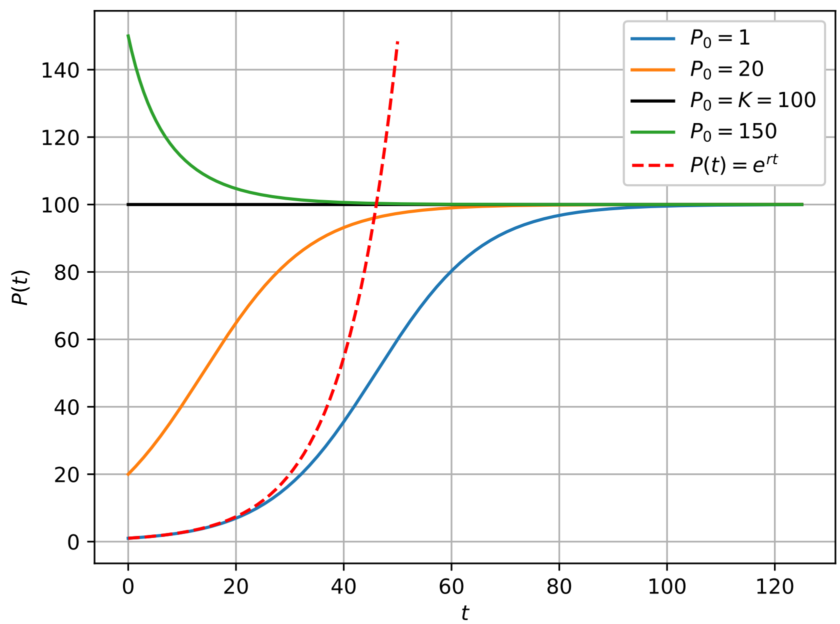


FIGURE 5.1. Graphs of some solutions $P(t)$ to the logistic equation with relative rate $r = 0.1$ and carrying capacity $K = 100$. Also shown for comparison is the exponential growth model e^{rt} with the same relative rate $r = 0.1$.

the study of oscillating physical systems. In the remainder of this introductory section, we make a few general remarks about some purely mathematical aspects of differential equations.

To expand upon what we saw by example above, a *differential equation* is an equation involving an unknown function $f(t)$ together with some of its derivatives $f'(t), f''(t), \ldots$. Here are some examples:

(i) $f'(t) = 2f(t)$

(ii) $f'(t) = -3f(t) + e^{-2t}$

(iii) $f'(t) = \sin(f(t))$

(iv) $f''(t) = -4f(t)$

(v) $f''(t) + f'(t) + f(t) = 0$

A *solution* to a differential equation is a particular function $f(t)$ that makes the equation true. For instance, the equation (i) above is an example of the population growth equation (with $r = 2$), and we have seen that for any constant C, the function $f(t) = Ce^{2t}$ is a solution:

$$f'(t) = (Ce^{2t})' = 2Ce^{2t} = 2f(t).$$

The next exercise asks you to verify some solutions to equation (iv).

EXERCISE 5.2. Consider the differential equation $f''(t) = -4f(t)$.

a) Check that the function $f(t) = \cos(2t)$ is a solution.
b) Check that the function $f(t) = 3\sin(2t)$ is a solution.
c) More generally, check that for all constants A, B, the following function is a solution:

$$f(t) = A\cos(2t) + B\sin(2t).$$

Problems 5.2 and 5.3 ask you to verify solutions to equations (ii) and (iii), and we will study equation (v) in Examples 5.3 and 5.4.

In the exponential and logistic population examples discussed earlier, we saw that we needed to specify an initial value $P(0) = P_0$ in order to fully determine a solution (see Figure 5.1) . In a similar way, in each of the examples listed above, there are actually infinitely many solutions to the differential equation:

(i) For any constant C, the function $f(t) = Ce^{2t}$ is a solution to $f'(t) = 2f(t)$. The constant C determines the *initial value* of the solution: $f(0) = Ce^{2\cdot 0} = C$.

(iv) For any constants A, B, the function $f(t) = A\cos(2t)+B\sin(2t)$ is a solution to $f''(t) = -4f(t)$. This time, the constants determine the initial values of the solution $f(t)$ and of the first derivative $f'(t)$:

$$\begin{aligned} f(0) &= A\cos(2\cdot 0) + B\sin(2\cdot 0) = A \\ f'(0) &= -2A\sin(2\cdot 0) + 2B\cos(2\cdot 0) = 2B. \end{aligned}$$

In general:

- When a differential equation involves only the first derivative of the unknown function $f(t)$ (as in (i–iii) above), we expect to obtain a family of solutions with one free constant, and making a specific choice for that constant allows us to obtain a solution satisfying a given *initial condition* for $f(t)$, often presented in the form $f(0) = a_0$ as in the population examples.

- When a differential equation involves the second derivative of the unknown function $f(t)$ (as in (iv–v) above), we expect to obtain a family of solutions with two free constants, and specific choices for these constants allow us to obtain a solution satisfying given *initial conditions* for $f(t)$ and $f'(t)$, often presented in the form $f(0) = a_0$ and $f'(0) = a_1$.
- An important *uniqueness theorem* guarantees that in these examples, there is only one solution of the equation satisfying the specified initial conditions.

EXERCISE 5.3. Find the solution $f(t)$ to the differential equation $f''(t) = -4f(t)$ satisfying the initial conditions $f(0) = 1$ and $f'(0) = -1$.

If you take a differential equations course, you will learn more about the existence of solutions to differential equations, and about the role of initial conditions in specifying a unique solution. You will also learn to identify different types of differential equations, and about special techniques for finding explicit solutions when they exist. Finally, you will see how differential equations arise in the social and physical sciences, and learn how to use mathematical properties of the solutions to gain insight into the original applied problems.

For all of the differential equations presented in this section, we have been able to write down explicit formulas for solutions involving only elementary functions. But in general, the solutions of differential equations do not have elementary formulas, and so we need other techniques to describe them—in the next section we will use power series to find solutions. But it is the differential equation itself that is the primary object of interest: as we saw in the population examples, a differential equation generally represents a story about a physical, social, or mathematical phenomenon. An *existence and uniqueness theorem* will generally guarantee that there is exactly one solution satisfying specified initial conditions, but we have no right to expect that this solution will be expressible in terms of familiar elementary functions. In fact, many so-called *special functions* in pure and applied mathematics are defined simply as solutions to certain differential equations—they come to our attention as solutions to interesting equations, and we give them a name in order to refer to them more easily. We will see an explicit example of such a function in Example 5.5 of the next section.

Key points from Section 5.1:

- What is a differential equation? (page 256)
- Verifying that a particular function f is a solution to a given differential equation (Exercise 5.1 and Exercise 5.2)
- What is an initial condition? (page 257, Exercise 5.3)

5.2. Series Solutions to Differential Equations

In this course, we have taken various real functions as known starting points (such as e^x, $\sin(x)$, and $\cos(x)$), and then used the good properties of power series to extend them into the complex domain. We present an alternative point of view in this section by showing how the complex functions (and thus the real functions as well), arise as power series solutions to some simple differential equations. We will also see how to find power series solutions to some differential equations that do not have elementary solutions.

Example 5.1. Suppose that we wish to find a complex differentiable function $f(z)$ satisfying the differential equation

$$f'(z) = f(z)$$

for all z near zero. Well, if such a function exists, then it is represented by its Maclaurin series, $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Substituting this expression into the differential equation yields:

$$\sum_{n=0}^{\infty} n a_n z^{n-1} = \sum_{n=0}^{\infty} a_n z^n.$$

For these two series to be equal, all of the coefficients must match, and so we find that

$$\begin{aligned} a_1 &= a_0 \\ 2a_2 &= a_1 \\ 3a_3 &= a_2 \\ &\vdots \\ na_n &= a_{n-1} \\ &\vdots \end{aligned}$$

It follows that we are free to choose the constant coefficient a_0, but then all the other coefficients are determined:

$$\begin{aligned} a_1 &= a_0 \\ a_2 &= \frac{a_1}{2} = \frac{a_0}{2} \\ a_3 &= \frac{a_2}{3} = \frac{a_0}{3 \cdot 2} \\ &\vdots \\ a_n &= \frac{a_{n-1}}{n} = \frac{a_0}{n!} \\ &\vdots \end{aligned}$$

So we see that

$$f(z) = \sum_{n=0}^{\infty} \frac{a_0}{n!} z^n = a_0 \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

The constant term a_0 specifies the value of $f(z)$ at the origin (the initial condition): $f(0) = a_0$. Note that our derivation reveals that $f(z)$ is the only solution satisfying the initial condition $f(0) = a_0$; we say that it is the *unique* solution.

You should recognize this power series as the complex exponential function, $f(z) = a_0 \exp(z)$, which now forces itself upon our attention as the solution of the simplest of all differential equations, $f'(z) = f(z)$. The next exercise asks you to generalize this example.

EXERCISE 5.4. Fix a complex number α and consider the differential equation $f'(z) = \alpha f(z)$. Mimic Example 5.1 to show that the unique solution satisfying the initial condition $f(0) = a_0$ is given by $f(z) = a_0 \exp(\alpha z)$.

EXAMPLE 5.2. Now consider the differential equation $f''(z) = f(z)$, where we are looking for a solution near $z = 0$. Once again, we introduce the Maclaurin series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ to get information about the unknown coefficients a_n:

$$\sum_{n=0}^{\infty} n(n-1)a_n z^{n-2} = \sum_{n=0}^{\infty} a_n z^n.$$

We find that

$$\begin{aligned} 2a_2 &= a_0 \\ 3 \cdot 2a_3 &= a_1 \\ 4 \cdot 3a_4 &= a_2 \\ &\vdots \\ n(n-1)a_n &= a_{n-2} \\ &\vdots \end{aligned}$$

In this case, we are free to choose a_0 and a_1, which then determine all other coefficients:

$$\begin{aligned} a_2 &= \frac{a_0}{2} \\ a_3 &= \frac{a_1}{3 \cdot 2} \\ a_4 &= \frac{a_2}{4 \cdot 3} = \frac{a_0}{4 \cdot 3 \cdot 2} \\ &\vdots \\ a_{2n} &= \frac{a_{2n-2}}{2n(2n-1)} = \frac{a_0}{(2n)!} \\ a_{2n+1} &= \frac{a_{2n-1}}{(2n+1)(2n)} = \frac{a_1}{(2n+1)!} \\ &\vdots \end{aligned}$$

So we find that

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} \left(\frac{a_0}{(2n)!} z^{2n} + \frac{a_1}{(2n+1)!} z^{2n+1} \right) \\ &= a_0 \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!} + a_1 \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}. \end{aligned}$$

The ratio test shows that these power series have radius of convergence $R = +\infty$, so that the solution $f(z)$ is defined for all complex z. We have $f(0) = a_0$ and $f'(0) = a_1$, so that these constants are the initial values of $f(z)$ and $f'(z)$. Note that taking $a_0 = a_1$ yields the solution $a_0 \exp(z)$ of the differential equation $f'(z) = f(z)$ from Example 5.1.

We have not previously encountered the two power series displayed above, but they certainly define interesting functions, as they are solutions to the simple differential equation $f''(z) = f(z)$. Moreover, these power series should remind you of the series for cosine and sine:

$$\cos(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}, \qquad \sin(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}.$$

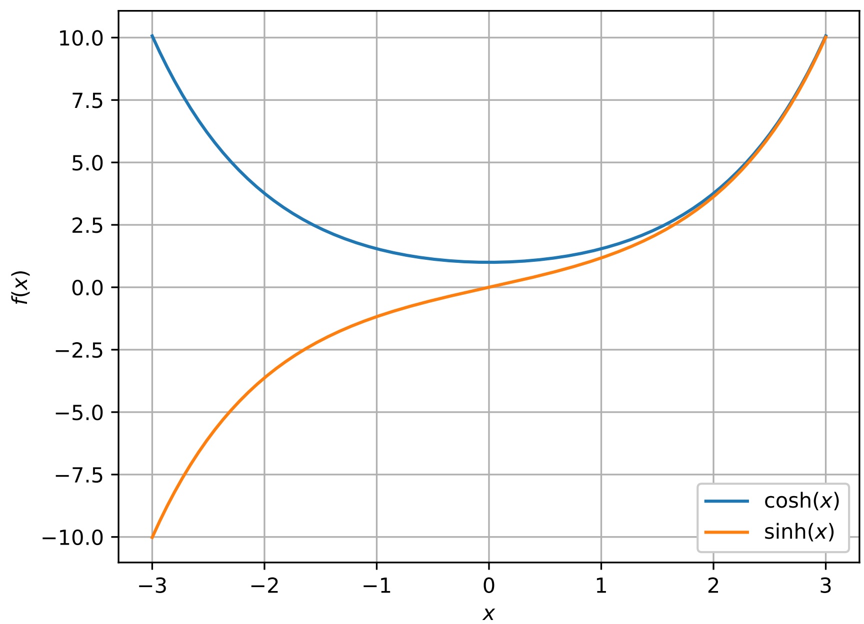


FIGURE 5.2. Graphs of the real functions $\cosh(x)$ and $\sinh(x)$.

So we give these new power series names, calling them the *hyperbolic cosine* and *hyperbolic sine* functions:

$$\cosh(z) = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}, \qquad \sinh(z) = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}.$$

In this notation, the general solution of $f''(z) = f(z)$ has the form

$$f(z) = a_0 \cosh(z) + a_1 \sinh(z).$$

Figure 5.2 displays graphs of the real versions of these functions, $\cosh(x)$ and $\sinh(x)$; see the optional Section 5.4 for their connection to hyperbolas and trigonometry.

EXERCISE 5.5. Now consider the equation $f''(z) = -f(z)$. Mimic Example 5.2 to show that the unique solution satisfying $f(0) = a_0$ and $f'(0) = a_1$ is given by

$$f(z) = a_0 \cos(z) + a_1 \sin(z).$$

Note that taking $a_1 = ia_0$ yields the solution $a_0 \exp(iz)$ to the differential equation $f'(z) = if(z)$.

EXAMPLE 5.3. Now consider the differential equation

$$f''(z) + f'(z) + f(z) = 0.$$

Introducing the Maclaurin series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and plugging into the differential equation yields:

$$\sum_{n=0}^{\infty} n(n-1)a_n z^{n-2} + \sum_{n=0}^{\infty} n a_n z^{n-1} + \sum_{n=0}^{\infty} a_n z^n = 0.$$

Now collect the coefficients of each power of z:

$$\begin{aligned}
2a_2 + a_1 + a_0 &= 0 \\
3 \cdot 2a_3 + 2a_2 + a_1 &= 0 \\
4 \cdot 3a_4 + 3a_3 + a_2 &= 0 \\
&\vdots \\
n(n-1)a_n + (n-1)a_{n-1} + a_{n-2} &= 0 \\
&\vdots
\end{aligned}$$

Solving for the highest coefficient in each line reveals

$$\begin{aligned}
a_2 &= -\frac{a_1}{2} - \frac{a_0}{2} \\
a_3 &= -\frac{a_2}{3} - \frac{a_1}{3 \cdot 2} = \frac{a_1}{3 \cdot 2} + \frac{a_0}{3 \cdot 2} - \frac{a_1}{3 \cdot 2} = \frac{a_0}{3 \cdot 2} \\
a_4 &= -\frac{a_3}{4} - \frac{a_2}{4 \cdot 3} = -\frac{a_0}{4!} + \frac{a_1}{4!} + \frac{a_0}{4!} = \frac{a_1}{4!} \\
a_5 &= -\frac{a_4}{5} - \frac{a_3}{5 \cdot 4} = -\frac{a_1}{5!} - \frac{a_0}{5!} \\
a_6 &= -\frac{a_5}{6} - \frac{a_4}{6 \cdot 5} = \frac{a_1}{6!} + \frac{a_0}{6!} - \frac{a_1}{6!} = \frac{a_0}{6!} \\
a_7 &= -\frac{a_6}{7} - \frac{a_5}{7 \cdot 6} = -\frac{a_0}{7!} + \frac{a_1}{7!} + \frac{a_0}{7!} = \frac{a_1}{7!} \\
&\vdots
\end{aligned}$$

We see that the pattern goes as follows:

$$\begin{aligned} a_{3n} &= \frac{a_0}{(3n)!} \\ a_{3n+1} &= \frac{a_1}{(3n+1)!} \\ a_{3n+2} &= -\frac{a_1}{(3n+2)!} - \frac{a_0}{(3n+2)!} \end{aligned}$$

So we wish to examine the power series

$$f(z) = \sum_{n=0}^{\infty} \left(\frac{a_0}{(3n)!} z^{3n} + \frac{a_1}{(3n+1)!} z^{3n+1} - \frac{a_0 + a_1}{(3n+2)!} z^{3n+2} \right).$$

An application of the ratio test shows that the series has radius of convergence $R = +\infty$, hence converges absolutely for all z. So, the series defines the unique solution $f(z)$ with $f(0) = a_0$ and $f'(0) = a_1$.

It is not apparent whether the power series solution from Example 5.3 can be expressed in terms of more familiar functions. This is the blessing and the curse of the series method for solving differential equations: it will often produce the power series representation of a solution even when no elementary formula exists, but when an elementary solution *does* exist, the series method may not reveal it. We illustrate by solving the differential equation of the previous example using a different method, that of *exponential trial solutions*.

EXAMPLE 5.4. Once again, we consider the differential equation

$$f''(z) + f'(z) + f(z) = 0.$$

This time, set $f(z) = \exp(\alpha z)$ for an unknown complex number α, and plug into the differential equation:

$$\begin{aligned} 0 &= f''(z) + f'(z) + f(z) \\ &= (\exp(\alpha z))'' + (\exp(\alpha z))' + \exp(\alpha z) \\ &= \alpha^2 \exp(\alpha z) + \alpha \exp(\alpha z) + \exp(\alpha z) \\ &= (\alpha^2 + \alpha + 1) \exp(\alpha z). \end{aligned}$$

It follows that we will have a solution if we choose α to be either root of the polynomial z^2+z+1. By the quadratic formula, these roots are

$$\alpha_1 = \frac{-1+\sqrt{3}\,i}{2} \qquad \alpha_2 = \frac{-1-\sqrt{3}\,i}{2}.$$

Notice that these roots are complex conjugates: $\alpha_2 = \overline{\alpha_1}$. So we find two solutions $f_1(z) = \exp(\alpha_1 z)$ and $f_2(z) = \exp(\alpha_2 z)$. The next exercise asks you to check that we may combine these to obtain more solutions:

EXERCISE 5.6. Show that for all constants A_1, A_2, the function $f(z)$ defined below is a solution to $f''(z) + f'(z) + f(z) = 0$:

$$f(z) = A_1 \exp(\alpha_1 z) + A_2 \exp(\alpha_2 z).$$

(Here, α_1 and α_2 are the roots of the polynomial $z^2 + z + 1$.)

If we want to match the initial conditions $f(0) = a_0$ and $f'(0) = a_1$, we just need to choose A_1 and A_2 appropriately:

$$a_0 = f(0) = A_1 \exp(\alpha_1 \cdot 0) + A_2 \exp(\alpha_2 \cdot 0) = A_1 + A_2$$

and

$$a_1 = f'(0) = A_1\alpha_1 \exp(\alpha_1 \cdot 0) + A_2\alpha_2 \exp(\alpha_2 \cdot 0) = A_1\alpha_1 + A_2\alpha_2.$$

EXERCISE 5.7. Solve the following pair of equations for A_1 and A_2:

$$\begin{aligned} A_1 + A_2 &= a_0 \\ A_1\alpha_1 + A_2\alpha_2 &= a_1. \end{aligned}$$

You should find that

$$A_1 = \frac{a_0\alpha_2 - a_1}{\alpha_2 - \alpha_1}, \qquad A_2 = \frac{a_0\alpha_1 - a_1}{\alpha_1 - \alpha_2}.$$

With these choices of A_1 and A_2, the function $f(z)$ is a solution to $f''(z) + f'(z) + f(z) = 0$ with the initial conditions $f(0) = a_0$ and $f'(0) = a_1$. But the power series solution from Example 5.3 is the unique such solution, so $f(z)$ must be equal to that power series.

Now let's look at the particular initial conditions $a_0 = 1$ and $a_1 = 0$. The coefficients are then

$$\begin{aligned} A_1 &= \frac{\alpha_2}{(\alpha_2 - \alpha_1)} = \frac{1+\sqrt{3}\,i}{2\sqrt{3}\,i} = \frac{1}{2} - \frac{i\sqrt{3}}{6} \\ A_2 &= 1 - A_1 = \frac{1}{2} + \frac{i\sqrt{3}}{6}. \end{aligned}$$

Notice that $A_2 = \overline{A_1}$.

Consider the first term of the solution $f(z)$:

$$\begin{aligned} A_1 \exp(\alpha_1 z) &= A_1 \exp\left(\frac{-z}{2} + \frac{i\sqrt{3}z}{2}\right) \\ &= A_1 \exp\left(\frac{-z}{2}\right) \exp\left(\frac{i\sqrt{3}z}{2}\right) \\ &= A_1 \exp\left(\frac{-z}{2}\right) \left(\cos\left(\frac{\sqrt{3}z}{2}\right) + i \sin\left(\frac{\sqrt{3}z}{2}\right)\right). \end{aligned}$$

And the second term:

$$\begin{aligned} A_2 \exp(\alpha_2 z) &= \overline{A_1} \exp\left(\frac{-z}{2} - \frac{i\sqrt{3}z}{2}\right) \\ &= \overline{A_1} \exp\left(\frac{-z}{2}\right) \exp\left(-\frac{i\sqrt{3}z}{2}\right) \\ &= \overline{A_1} \exp\left(\frac{-z}{2}\right) \left(\cos\left(\frac{\sqrt{3}z}{2}\right) - i \sin\left(\frac{\sqrt{3}z}{2}\right)\right). \end{aligned}$$

Adding these together yields the solution $f(z)$:

$$\begin{aligned} f(z) &= \exp\left(\frac{-z}{2}\right) \left(A_1 \exp\left(\frac{i\sqrt{3}z}{2}\right) + \overline{A_1} \exp\left(-\frac{i\sqrt{3}z}{2}\right)\right) \\ &= \exp\left(\frac{-z}{2}\right) \left(\cos\left(\frac{\sqrt{3}z}{2}\right) + \frac{\sqrt{3}}{3} \sin\left(\frac{\sqrt{3}z}{2}\right)\right). \end{aligned}$$

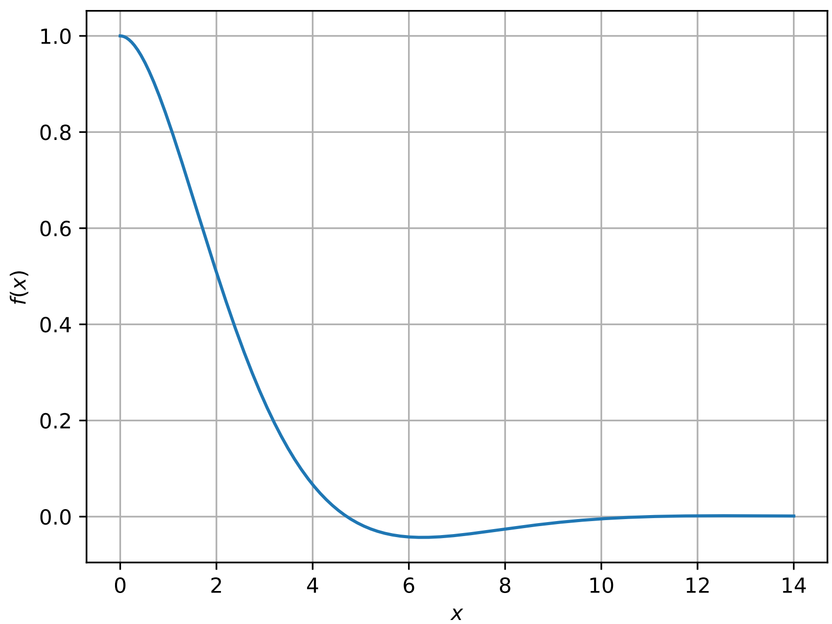


FIGURE 5.3. Plot showing the graph of the real solution $f(x) = e^{-x/2}\left(\cos\left(\frac{\sqrt{3}x}{2}\right) + \frac{\sqrt{3}}{3}\sin\left(\frac{\sqrt{3}x}{2}\right)\right)$ to the differential equation $f''(x) + f'(x) + f(x) = 0$.

Figure 5.3 shows the graph of the corresponding real solution obtained by restricting to real inputs $z = x$. In Section 5.3 we will discuss an interpretation of this solution in terms of oscillating physical systems.

If you take a differential equations course, you will learn more techniques for finding solutions of differential equations in terms of familiar elementary functions. But as mentioned at the end of Section 5.1, often the solutions are not expressible in terms of elementary functions, and in those cases the series method is invaluable. The next example provides a prominent example.

EXAMPLE 5.5. Consider the differential equation

$$z^2 f''(z) + zf'(z) + z^2 f(z) = 0.$$

Note the difference from our earlier examples: the coefficients of this differential equation (z^2, z, and z^2) are functions of z, rather than constants. Nevertheless, we again search for a power series solution near $z = 0$, and so we introduce a Maclaurin series $f(z) = \sum_{n=0}^{\infty} a_n z^n$, with currently unknown coefficients a_n. We begin by computing each term of the differential equation separately:

$$\begin{aligned} z^2 f''(z) &= z^2 \sum_{n=0}^{\infty} n(n-1)a_n z^{n-2} = \sum_{n=0}^{\infty} n(n-1)a_n z^n \\ z f'(z) &= z \sum_{n=0}^{\infty} n a_n z^{n-1} = \sum_{n=0}^{\infty} n a_n z^n \\ z^2 f(z) &= z^2 \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n z^{n+2}. \end{aligned}$$

Now add these together to regain the differential equation:

$$\begin{aligned} 0 &= z^2 f''(z) + z f'(z) + z^2 f(z) \\ &= \sum_{n=0}^{\infty} n(n-1)a_n z^n + \sum_{n=0}^{\infty} n a_n z^n + \sum_{n=0}^{\infty} a_n z^{n+2} \\ &= a_1 z + \sum_{n=2}^{\infty} (n(n-1)a_n + n a_n + a_{n-2}) z^n \\ &= a_1 z + \sum_{n=2}^{\infty} (n^2 a_n + a_{n-2}) z^n. \end{aligned}$$

It follows that

$$\begin{aligned} a_1 &= 0 \\ 2^2 a_2 + a_0 &= 0 \\ 3^2 a_3 + a_1 &= 0 \\ 4^2 a_4 + a_2 &= 0 \\ &\vdots \\ n^2 a_n + a_{n-2} &= 0 \\ &\vdots \end{aligned}$$

Solving for the highest index term in each line:

$$\begin{aligned}
a_1 &= 0\\
a_2 &= -\frac{a_0}{2^2}\\
a_3 &= -\frac{a_1}{3^2} = 0\\
a_4 &= -\frac{a_2}{4^2} = \frac{a_0}{(4\cdot 2)^2}\\
a_5 &= -\frac{a_3}{5^2} = 0\\
a_6 &= -\frac{a_4}{6^2} = -\frac{a_0}{(6\cdot 4\cdot 2)^2}\\
&\vdots
\end{aligned}$$

Observe the developing pattern:

- all odd terms are zero: $a_{2n+1} = 0$;
- the even terms alternate in sign;
- the denominator of a_{2n} is the square of the product of all even numbers between 2 and $2n$:

  $$a_{2n} = (-1)^n \frac{a_0}{(2n\cdot(2n-2)\cdots 4\cdot 2)^2}.$$

  We can rewrite this product of even numbers as follows:

  $$2n\cdot(2n-2)\cdots 4\cdot 2 = 2^n(n\cdot(n-1)\cdots 2\cdot 1) = 2^n\cdot n!.$$

  Thus, the formula for the coefficient a_{2n} may be written as follows:

  $$a_{2n} = (-1)^n \frac{a_0}{2^{2n}(n!)^2}.$$

So the power series solution is

$$f(z) = a_0 \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{2^{2n}(n!)^2} = a_0\left(1 - \frac{z^2}{4} + \frac{z^4}{64} - \frac{z^6}{2304} + \cdots\right).$$

Now find the radius of convergence:

$$\frac{|a_{2n+2}|}{|a_{2n}|} = \frac{2^{2n}(n!)^2}{2^{2n+2}((n+1)!)^2} = \frac{1}{4(n+1)^2} \to L = 0.$$

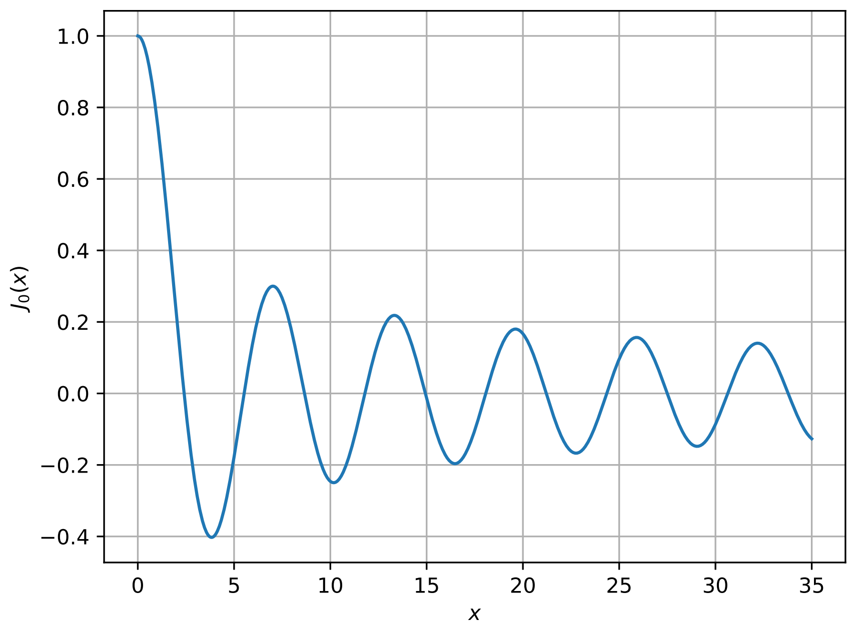


FIGURE 5.4. Graph of the real Bessel function $J_0(x)$.

It follows that $R = +\infty$, and the series converges for all complex z. If we take the solution satisfying the initial condition $f(0) = a_0 = 1$, we obtain the *Bessel function of the first kind of order zero*:

$$J_0(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{2^{2n}(n!)^2}.$$

Figure 5.4 shows the graph of the corresponding real function $J_0(x)$.

Note that the series method has produced a family of solutions with only one free constant a_0 rather than the two free constants a_0 and a_1 that we may have expected based on the fact that the differential equation involves the second derivative $f''(z)$. In fact, there is a second family of solutions, but they are not defined at $z = 0$, and hence our series method did not discover them (these are *Bessel functions of the second kind*).

The differential equation $z^2 f''(z) + zf'(z) + z^2 f(z) = 0$ is known as *Bessel's differential equation of order zero*. More generally, for any nonnegative integer $k \geq 0$, *Bessel's differential equation of order* k is

$$z^2 f''(z) + z f'(z) + (z^2 - k^2) f(z) = 0.$$

As in the case $k = 0$, application of the series method to this equation yields a family of solutions with one complex parameter a_k, and an appropriate choice for a_k yields $J_k(z)$, the *Bessel function of the first kind of order* k (see Problems 5.6 and 5.7).

The point of this example is the following: the Bessel functions $J_k(z)$ arise from the consideration of some simple differential equations, and they appear in many areas of mathematics and physics. They are defined by complex power series, and it is simply not possible to express them in terms of other more familiar and elementary functions.

Key points from Section 5.2:

- Series method for solving differential equations (Examples 5.1, 5.2, 5.3, 5.5)
- Method of exponential trial solutions (Example 5.4)

5.3. Optional: Oscillators and the Complex Exponential

In this section, we illustrate how some of the differential equations from the previous section arise in the important context of oscillating physical systems. As explained at the end of this section, the model that we are about to introduce is ubiquitous, because it serves as a good approximation to most physical systems near stable equilibria (which is often where we want to study them). To begin, consider a spring attached at one end to the wall, with the other end attached to a moveable block of mass m (See Figure 5.5). The mass is resting on the floor, which is covered with oil. If you pull the block away from the wall and then let go, the stretched spring will pull the block back toward the wall, while the friction with the oily floor tends to impede the block's motion. Once the block passes its original resting place the spring will compress, pushing on the block until it stops moving toward the wall and starts moving away under the force of the now decompressing spring. We expect the block to continue moving back and forth, with the oscillations getting smaller and smaller due to the friction with the oily floor. In this section, we will write down a mathematical model of this setup in the form of a differential equation, and the solutions will provide the possible motions of the block.

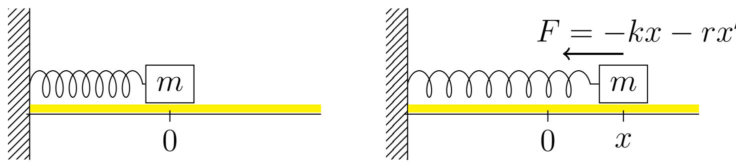


FIGURE 5.5. A block of mass m connected to the wall by a spring with spring constant k. The floor is covered with oil that has a friction coefficient r.

First, we must set up the notation: let $x(t)$ denote the position of the block at time t, with $x = 0$ indicating the resting position of the block, in which the spring is at its natural length, neither compressed nor stretched. Increasing values of x indicate motion away from the wall. Recall that the *velocity* of the block is given by the derivative $x'(t)$, while the *acceleration* is given by the derivative of velocity, or the second derivative of position: $x''(t)$. In the setup described above, the block experiences two types of forces:

- a force s due to the stretched/compressed spring: we model this as a linear function of position $s(x) = -kx$ for a positive *spring constant* k. The negative sign indicates that when $x > 0$, the stretched spring pulls the block back toward the wall, while when $x < 0$, the compressed spring pushes the block away from the wall. Larger values of k correspond to stiffer springs.
- a friction force f due to the oily floor: we model this as a linear function of velocity $f(x') = -rx'$ for a positive *friction coefficient* r. The negative sign indicates that the friction always acts in the opposite direction of the block's velocity, tending

to resist its motion. Larger values of r correspond to more viscous oil.

Now we apply Newton's Second Law of Motion, which states that if our block feels any type of force F, it responds by accelerating in the direction of the force, with the block's mass acting as the constant of proportionality:

$$F = mx''.$$

In our setup, the total force $F = s + f$ is the sum of the spring force and the friction force, and we find that

$$mx'' = F = s(x) + f(x') = -kx - rx'.$$

Rearranging, we obtain the following differential equation for the position function $x(t)$:

$$x''(t) + \frac{r}{m}x'(t) + \frac{k}{m}x(t) = 0.$$

For convenience, we now assume that our block has mass $m = 1$ (which just amounts to adjusting our units), so we can write our equation as

$$x''(t) + rx'(t) + kx(t) = 0.$$

Even though we are looking for a real function $x(t)$, we begin by plugging in the complex trial solution $z(t) = \exp(\alpha t)$, using an unknown complex constant α. We proceed as in Example 5.4:

$$\begin{aligned}
0 &= z''(t) + rz'(t) + kz(t) \\
&= (\exp(\alpha t))'' + r(\exp(\alpha t))' + k\exp(\alpha t) \\
&= \alpha^2\exp(\alpha t) + r\alpha\exp(\alpha t) + k\exp(\alpha t) \\
&= (\alpha^2 + r\alpha + k)\exp(\alpha z).
\end{aligned}$$

We find that we will have a solution provided that α is a root of the quadratic polynomial $w^2 + rw + k$. By the quadratic formula, the roots are

$$\alpha_1 = \frac{-r + \sqrt{r^2 - 4k}}{2} \quad \text{and} \quad \alpha_2 = \frac{-r - \sqrt{r^2 - 4k}}{2}.$$

There are several cases to consider:

- ($d = r^2 - 4k > 0$) This is the "viscous oil and weak spring" regime, where $\alpha_2 < \alpha_1 < 0$ are negative real numbers.
- ($d = r^2 - 4k < 0$) This is the "non-viscous oil and strong spring" regime, where $\alpha_1 \neq \alpha_2$ are complex conjugate complex numbers.
- ($d = r^2 - 4k = 0$) This is the "critical damping" regime where there is only a single real root $\alpha_1 = \alpha_2 = -r/2$.

We begin by assuming $r^2 \neq 4k$, so we have two distinct roots $\alpha_1 \neq \alpha_2$. Then the most general complex solution looks like

$$z(t) = A_1 \exp(\alpha_1 t) + A_2 \exp(\alpha_2 t),$$

where the coefficients must be chosen to match the desired initial conditions $z(0) = a_0$ and $z'(0) = a_1$. As in Example 5.4, we will consider the initial conditions $z(0) = 1$ and $z'(0) = 0$, which correspond to stretching the block a unit distance away from its resting location, holding the block steady, and then letting go at time $t = 0$, so that the block starts with an initial velocity of zero. To achieve these initial conditions, we must have (Problem 5.5)

$$A_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1} = \frac{r + \sqrt{d}}{2\sqrt{d}} = \frac{1}{2} + \frac{r\sqrt{d}}{2d}.$$

and

$$A_2 = 1 - A_1 = \frac{1}{2} - \frac{r\sqrt{d}}{2d}.$$

In the $d > 0$ regime, all of these quantities are real, and we see immediately that the real solution is

$$x(t) = A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t}.$$

Both terms are decaying exponentials, with the second having a faster rate of decay. Figure 5.6 shows the graph of this solution for the case $r = 3$ and $k = 1$. There is no oscillation: the oil is so viscous compared to the spring strength that the block just slides back to zero. This phenomenon is called *overdamping*.

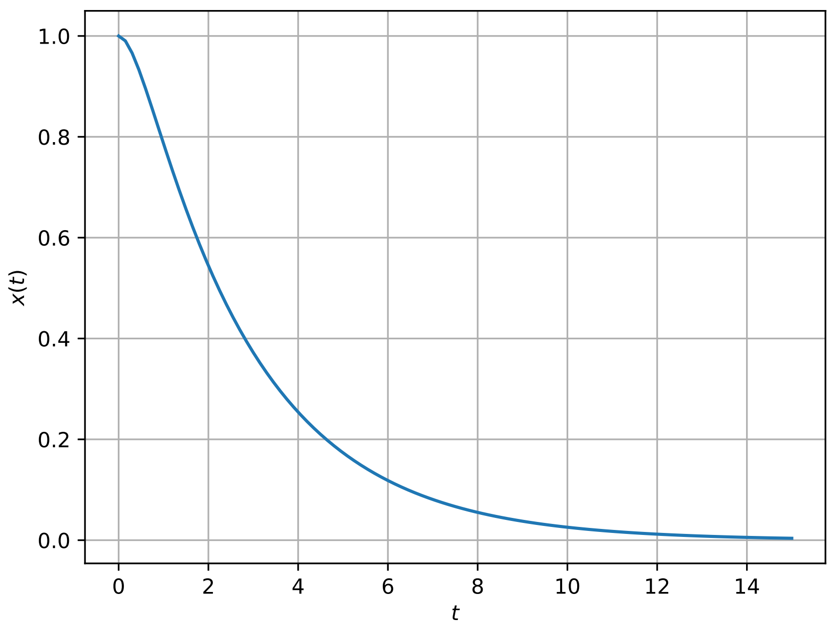


FIGURE 5.6. Graph of an overdamped oscillator, with $r = 3$ and $k = 1$.

Now consider the $d < 0$ regime, where all quantities are complex. Generalizing Example 5.4 (see Problem 5.5), the real solution is

$$x(t) = \exp\left(\frac{-rt}{2}\right)\left(\cos\left(\frac{\sqrt{|d|}\,t}{2}\right) + \frac{r\sqrt{|d|}}{|d|}\sin\left(\frac{\sqrt{|d|}\,t}{2}\right)\right).$$

Figure 5.7 shows the graph of this solution for $r = 1$ and $k = 10$. Here we see the expected oscillation, decaying over time due to the friction with the oily floor. This phenomenon is called *underdamping*.

It remains to consider the case of *critical damping*, were $d = 0$ and there is a single real root $\alpha_1 = \alpha_2 = -r/2$. In this case, we find only one family of solutions $A_1 \exp(-\frac{rt}{2})$. But we cannot choose the constant A_1

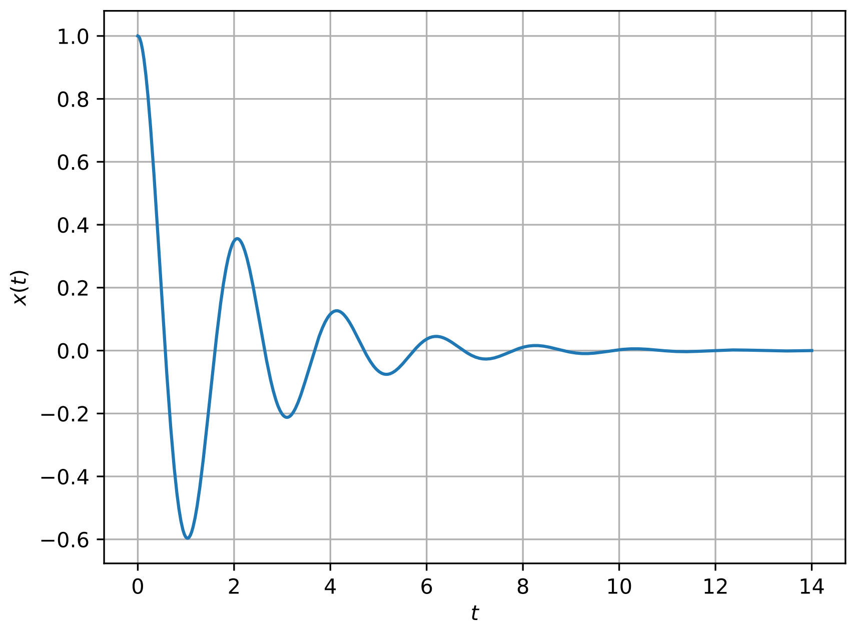


FIGURE 5.7. Graph of an underdamped oscillator, with $r = 1$ and $k = 10$.

so as to achieve the initial conditions $x(0) = 1$ and $x'(0) = 0$. Luckily, we can check that $t\exp(-\frac{rt}{2})$ is another solution.

EXERCISE 5.8. Show by direct computation that $x(t) = t\exp(-\frac{rt}{2})$ is a solution to the critically damped oscillator differential equation

$$x''(t) + rx'(t) + \frac{r^2}{4}x(t) = 0.$$

So now the general solution looks like

$$x(t) = A_1 \exp\left(-\frac{rt}{2}\right) + A_2 t \exp\left(-\frac{rt}{2}\right),$$

and we wish to choose A_1 and A_2 to satisfy the initial conditions. Setting $1 = x(0) = A_1$ determines the first coefficient, and then

$$0 = x'(0) = -\frac{r}{2} + A_2$$

implies that $A_2 = \frac{r}{2}$. The corresponding solution is thus

$$x(t) = e^{-\frac{rt}{2}} + \frac{rt}{2}e^{-\frac{rt}{2}} = e^{-\frac{rt}{2}}\left(1 + \frac{rt}{2}\right).$$

The graph of this solution for $r = 2$ and $k = 1$ is shown in Figure 5.8. The term *critically damped* refers to the fact that this exact combination of friction r and spring-stiffness k results in the fastest possible damping of the oscillator: any deviation from the critical point $r^2 = 4k$ will yield either oscillations (underdamped) or a slower slide across the oily floor (overdamped). Because of this optimal behavior for damping unwanted vibrations, the phenomenon of critical damping is important for many engineering applications.

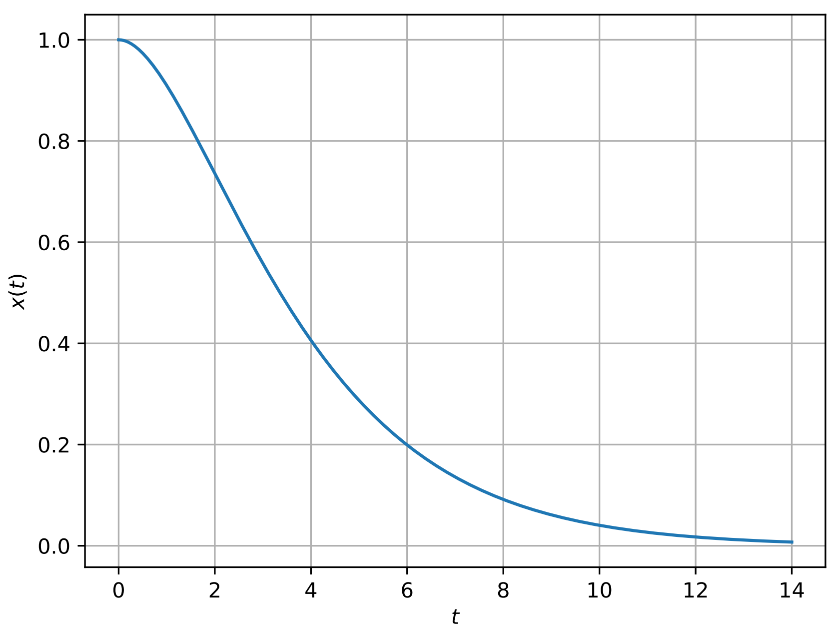


FIGURE 5.8. Graph of a critically damped oscillator, with $r = 2$ and $k = 1$.

Now, you may be wondering how important these oscillator differential equations can possibly be. Indeed, you are not likely to encounter many physical systems manufactured out of springs and oil as described above. But miraculously, the same differential equations govern the behavior of many physical systems that seem initially to have nothing in common with springs and oil. As a prominent example, when those of you studying physics take an electronics course, you will discover that these same equations describe the behavior of RLC circuits, consisting of a resistor, an inductor, and a capacitor connected in series or in parallel.

An important special case of the oscillator equation comes from setting the friction coefficient $r = 0$; this corresponds to a frictionless floor (think of an air-hockey table). In this case, the differential equation becomes (reinstating the mass m, which we previously set to $m = 1$):

$$z''(t) + \frac{k}{m} z(t) = 0.$$

This *undamped oscillator* is in the underdamped regime, with real solution (still using the initial conditions $x(0) = 1$ and $x'(0) = 0$):

$$x(t) = \cos(\omega t) \qquad \text{with } \omega = \sqrt{\frac{k}{m}}.$$

Figure 5.9 shows the graph of this familiar solution: there is no decay, and the oscillations repeat with period $2\pi/\omega = 2\pi\sqrt{\frac{m}{k}}$. This behavior is called *simple harmonic motion*. In particular, we see that increasing the mass leads to slower oscillations, while stiffening the spring leads to quicker ones.

The simple harmonic oscillator equation $x''(t) + \omega^2 x(t) = 0$ and the resulting simple harmonic motion are of universal importance in physics, and we can understand why by applying the theory of Taylor polynomials to a general 1-dimensional physical problem. Here is the setup: suppose that we have a particle of mass m moving along the x-axis due to the presence of a force $F(x)$ that varies with the location of the particle. By Newton's Second Law, the particle experiences an acceleration given by

$$x''(t) = \frac{1}{m} F(x(t)).$$

At this point, the force $F(x)$ is quite general, and we want to keep it that way. But let us at least assume that $F(x)$ is a continuous function,

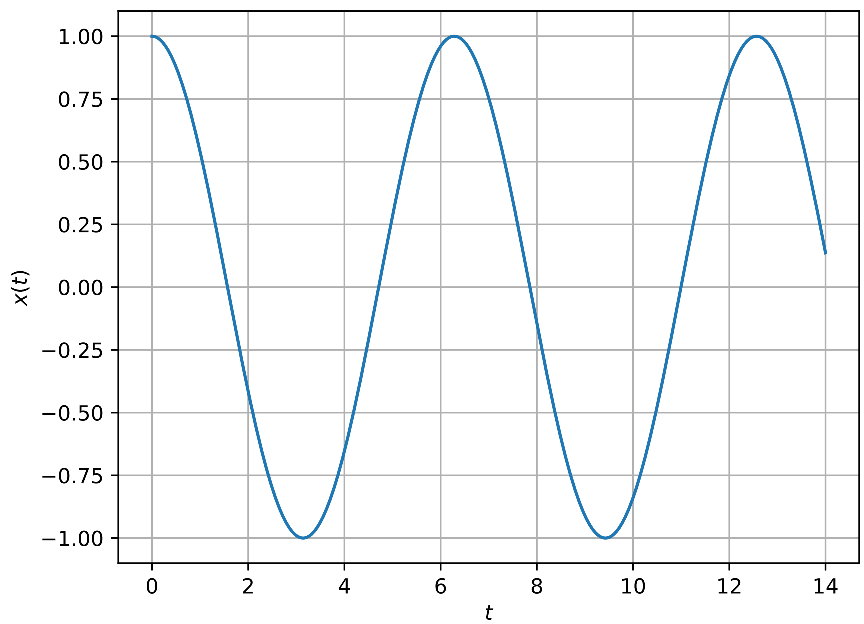


FIGURE 5.9. Graph of undamped oscillator, with constants $k = m = 1$.

which means that it has an antiderivative. In fact, we will introduce an antiderivative for $-F(x)$ and call it the *potential function* $V(x)$:

$$V'(x) = -F(x).$$

(Note that, given the force $F(x)$, the potential $V(x)$ is determined only up to an additive constant.)

Now step back for a minute, and think about the stability of the world around you: there are many small objects in your immediate vicinity, and most of them don't seem to be moving very much. Moreover, if they do move a bit, they often tend to come back to where they started, perhaps rocking or rolling around for a while. Think, for example, of a grape sitting at the bottom of an otherwise empty fruit bowl. Of course, if you slam your hand into the grape, it may fly out of the bowl. But if you only nudge it a bit, then it will roll around only slightly, the walls of the bowl combined with gravity forcing it back

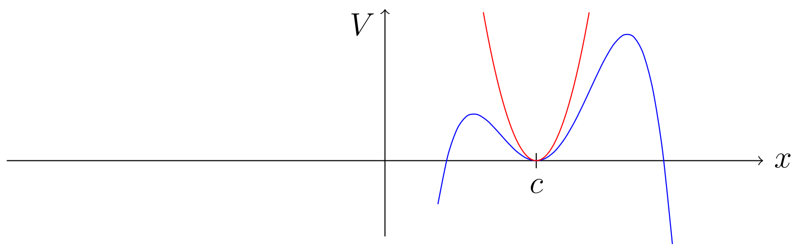


FIGURE 5.10. The point c is a stable equilibrium for the potential function $V(x)$ (shown in blue), with quadratic Taylor polynomial $T_2(x)$ (shown in red).

toward the center. So: we are often interested in studying particles at locations where the force on them is zero, and near which they experience small forces that push them back toward their starting point . . . we call such points *stable equilibria*. Let's think about what the graph of the potential function $V(x)$ looks like near a stable equilibrium at $x = c$ (see Figure 5.10):

- First of all, we can subtract a constant from the potential to ensure that $V(c) = 0$.
- There is no force at $x = c$, so $0 = F(c) = -V'(c)$, and the graph of $V(x)$ has a horizontal tangent line at $x = c$.
- The forces near $x = c$ tend to push the particle back toward c, so $0 < F(x) = -V'(x)$ for $x < c$ and $0 > F(x) = -V'(x)$ for $x > c$. Thus we see that $V'(x)$ is negative to the left of c and positive to the right of c. By the first derivative test, the function $V(x)$ has a local minimum at $x = c$.
- If we also assume that the second derivative of V exists and is continuous, then we can apply the second derivative test to conclude that $V''(c) \geq 0$.

Now consider the quadratic Taylor polynomial $T_2(x)$ for the potential $V(x)$ at c:

$$T_2(x) = V(c) + V'(c)(x - c) + \frac{V''(c)}{2}(x - c)^2 = \frac{\omega^2}{2}(x - c)^2,$$

where we have introduced the nonnegative constant $\omega = \sqrt{V''(c)}$. Recall that this is the best quadratic approximation to the function $V(x)$ near the stable equilibrium $x = c$.

Depending on the exact nature of the potential $V(x)$, the Taylor polynomial $T_2(x)$ may or may not be a good approximation near $x = c$. But in most cases, the potential $V(x)$ is actually equal to its Taylor series near c, and moreover $V''(c)$ is nonzero (so that $\omega > 0$). In this case, the quadratic approximation is often quite good, in which case we can approximate the particle's motion by using the potential

$$\widetilde{V}(x) = T_2(x) = \frac{\omega^2}{2}(x - c)^2,$$

and this yields an approximate force $\widetilde{F}(x) = -\widetilde{V}'(x) = -\omega^2(x - c)$. Now return to Newton's Second Law with these approximations:

$$x''(t) = \frac{1}{m}\widetilde{F}(x(t)) = -\omega^2(x(t) - c).$$

Making the change of variable $u = x - c$ (which just amounts to translating the origin), we find that $u''(t) = x''(t)$, and so

$$u''(t) + \omega^2 u(t) = 0,$$

which is the simple harmonic oscillator equation. This means that the solution $u(t)$ describes simple harmonic motion about the origin, or equivalently, that $x(t) = u(t) + c$ describes simple harmonic motion about the stable equilibrium $x = c$. In summary: *to second-order approximation, all particles exhibit simple harmonic motion near a stable equilibrium.* This wonderful fact accounts for the ubiquity of the simple harmonic oscillator, and makes it the most important piece of physics that you will ever learn.

Key points from Section 5.3:

- The damped oscillator equation $x'' + \frac{r}{m}x' + \frac{k}{m}x = 0$ (page 274)
- Overdamping, underdamping, and critical damping (pages 274–278)
- The ubiquity of simple harmonic motion (pages 279–282)

5.4. Optional: Hyperbolic Trigonometry

In Example 5.2, we introduced two power series as solutions to the differential equation $f''(z) = f(z)$, calling them the *hyperbolic cosine* and *hyperbolic sine* functions:

$$\cosh(z) = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}, \qquad \sinh(z) = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}.$$

In this section, we derive some properties of these functions and explain their names.

Exercise 5.9. Use the series definitions to show that

$$(\sinh(z))' = \cosh(z) \quad \text{and} \quad (\cosh(z))' = \sinh(z).$$

Remark 5.6. You should compare the previous exercise to the situation for the circular trigonometric functions:

$$(\sin(z))' = \cos(z) \quad \text{and} \quad (\cos(z))' = -\sin(z).$$

The next exercise shows that $\cosh(z)$ and $\sinh(z)$ are the even and odd parts of the complex exponential function.

Exercise 5.10.

a) Using the power series definitions, show that

$$\exp(z) = \cosh(z) + \sinh(z).$$

b) Show that $\cosh(z)$ is an even function and $\sinh(z)$ is an odd function:

$$\cosh(-z) = \cosh(z), \qquad \sinh(-z) = -\sinh(z).$$

c) Use parts a) and b) to show that

$$\cosh(z) = \frac{\exp(z) + \exp(-z)}{2} \qquad \text{and} \qquad \sinh(z) = \frac{\exp(z) - \exp(-z)}{2}.$$

Remark 5.7. You should compare the previous exercise to the following situation for the circular trigonometric functions $\cos(z)$ and $\sin(z)$:

$$\begin{aligned} \exp(iz) &= \cos(z) + i\sin(z) \\ \cos(z) &= \frac{\exp(iz) + \exp(-iz)}{2} \\ i\sin(z) &= \frac{\exp(iz) - \exp(-iz)}{2}. \end{aligned}$$

These relations reveal that $\cos(z)$ and $i\sin(z)$ are the even and odd parts of the function $\exp(iz)$. Restricting to real values $z = \theta$, we obtain the more familiar fact that $\cos(\theta)$ and $\sin(\theta)$ are the real and imaginary parts of $\exp(i\theta)$:

$$\begin{aligned} \exp(i\theta) &= \cos(\theta) + i\sin(\theta) \\ \cos(\theta) &= \frac{\exp(i\theta) + \exp(-i\theta)}{2} = \mathrm{Re}(\exp(i\theta)) \\ \sin(\theta) &= \frac{\exp(i\theta) - \exp(-i\theta)}{2i} = \mathrm{Im}(\exp(i\theta)). \end{aligned}$$

Just as we obtained circular trigonometric identities by taking powers of de Moivre's formula $\exp(i\theta) = \cos(\theta) + i\sin(\theta)$, we obtain identities for the hyperbolic functions $\cosh(z)$ and $\sinh(z)$ by taking powers of the formula $\exp(z) = \cosh(z) + \sinh(z)$. For instance:

$$\begin{aligned} \exp(2z) &= \big(\exp(z)\big)^2 \\ &= (\cosh(z) + \sinh(z))^2 \\ &= \cosh^2(z) + 2\cosh(z)\sinh(z) + \sinh^2(z). \end{aligned}$$

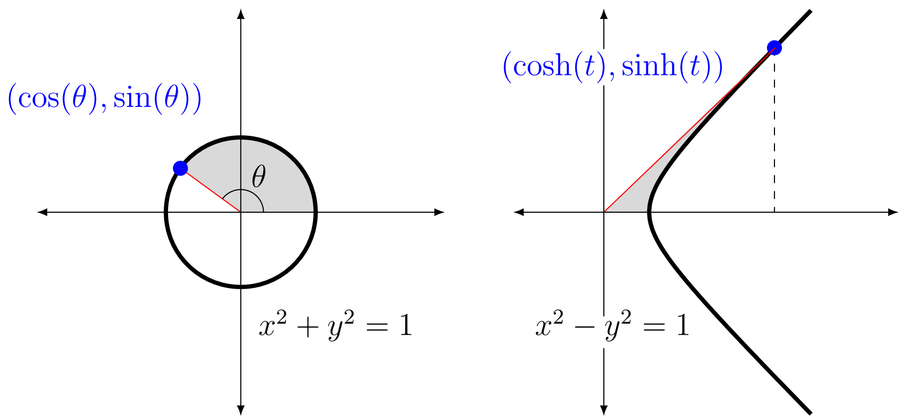


FIGURE 5.11. The left hand side shows the parametrization of the unit circle using the circular trigonometric functions $(\cos(\theta), \sin(\theta))$. The right hand side shows the parametrization of the unit hyperbola using the hyperbolic trigonometric functions $(\cosh(t), \sinh(t))$. In each case, the area of the shaded region is equal to one-half of the parameter (θ or t).

Now take the even and odd parts of both sides:

$$\begin{aligned}\cosh(2z) &= \cosh^2(z) + \sinh^2(z) \\ \sinh(2z) &= 2\cosh(z)\sinh(z).\end{aligned}$$

To understand the term *hyperbolic* as applied to these functions, restrict attention to real values $z = t$. Then we have

$$1 = e^t e^{-t} = (\cosh(t) + \sinh(t))(\cosh(t) - \sinh(t)) = \cosh^2(t) - \sinh^2(t).$$

This means that as t varies, the point $(\cosh(t), \sinh(t))$ traces out the right half of the unit hyperbola $x^2 - y^2 = 1$ (see right side of Figure 5.11). Compare this to the way that the point $(\cos(\theta), \sin(\theta))$ traces out the unit circle $x^2 + y^2 = 1$.

It is interesting to think more carefully about these parametrizations of the circle and the hyperbola.

EXERCISE 5.11. Convince yourself that the area of the shaded region on the left side of Figure 5.11 is equal to $\theta/2$. Thus, the parameter θ in $[0, 2\pi)$ for the unit circle may be interpreted as twice the area of the shaded region swept out by the red radius pointing to $(\cos(\theta), \sin(\theta))$.

We wish to show that the parameter t for the hyperbola also has an interpretation as an area. Consider the shaded region on the right in Figure 5.11, bounded by the x-axis, the hyperbola, and the red line connecting the origin to the point $(\cosh(t), \sinh(t))$. Let $A(t)$ denote the area of this shaded region. Also consider the triangle with vertices $(0,0), (\cosh(t), 0)$, and $(\cosh(t), \sinh(t))$. Assuming that $t > 0$ as in the figure, the area of the triangle is $\frac{1}{2}\cosh(t)\sinh(t)$. Moreover, the area $A(t)$ is the difference between the triangular area and the area under the hyperbola on the interval $[1, \cosh(t)]$:

$$A(t) = \frac{1}{2}\cosh(t)\sinh(t) - \int_1^{\cosh(t)} \sqrt{x^2-1}dx.$$

To compute the integral, we will make use of the following identities established earlier:

$$\begin{aligned}
\cosh^2(v) - 1 &= \sinh^2(v) \\
\cosh(2v) &= \cosh^2(v) + \sinh^2(v) = 1 + 2\sinh^2(v).
\end{aligned}$$

Now make the substitution $x = \cosh(v)$, so that $dx = \sinh(v)dv$:

$$\begin{aligned}
\int_1^{\cosh(t)} \sqrt{x^2-1}dx &= \int_0^t \sqrt{\left(\cosh^2(v) - 1\right)} \sinh(v)dv \\
&= \int_0^t \sinh^2(v)dv \\
&= \frac{1}{2}\int_0^t (\cosh(2v) - 1)dv \\
&= \left(\frac{\sinh(2v)}{4} - \frac{v}{2}\right)\Big|_0^t \\
&= \frac{\sinh(2t)}{4} - \frac{t}{2}.
\end{aligned}$$

Using the identity $\sinh(2t) = 2\cosh(t)\sinh(t)$, we find that

$$\begin{aligned} A(t) &= \frac{1}{2}\cosh(t)\sinh(t) - \frac{\sinh(2t)}{4} + \frac{t}{2} \\ &= \frac{t}{2}. \end{aligned}$$

Hence, the parameter $t = 2A(t)$ is twice the area of the shaded region in Figure 5.11 swept out by the red line pointing from the origin to the point $(\cosh(t), \sinh(t))$.

The hyperbolic trigonometric functions $\cosh(z)$ and $\sinh(z)$ have many applications throughout mathematics, physics, and engineering. We mention just two:

- They appear in the description of the Lorentz transformations of special relativity, serving to relate the spacetime measurements of two observers in uniform relative motion;
- For any constants $A, B > 0$, the graph of $A\cosh(x/B)$ is called a *weighted catenary curve*. When $A = B$, the curve is an *ordinary catenary* and describes the shape of an idealized hanging chain supported only at its two ends. For this reason, weighted catenaries are commonly used in engineering and architectural applications. A prominent example is the St. Louis Gateway Arch (Figure 5.12).

Key points from Section 5.4:

- The hyperbolic trigonometric functions $\cosh(z)$ and $\sinh(z)$ (page 283)
- The identity $\cosh^2(t) - \sinh^2(t) = 1$ and parametrization of the unit hyperbola $x^2 - y^2 = 1$ (Figure 5.11)

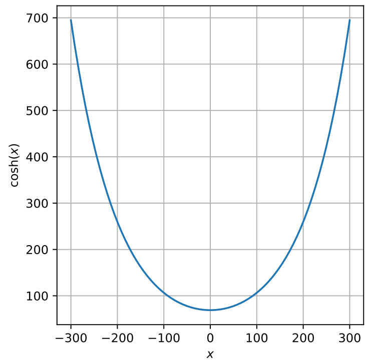


FIGURE 5.12. Top: Graph of the weighted catenary $69\cosh(x/100)$; Bottom: "Gateway Arch & St. Louis MO Riverfront at Dawn" by Parker Botanical is licensed under CC BY-SA 3.0.

5.5. In-text Exercises

This section collects the in-text exercises that you should have worked on while reading the chapter.

EXERCISE 5.1 Consider the logistic differential equation with relative rate $r = 0.1$ and carrying capacity $K = 100$:

$$P'(t) = 0.1P(t) - \frac{0.1}{100}(P(t))^2$$

a) By direct computation, verify that the following function is a solution satisfying the initial condition $P(0) = P_0$:

$$P(t) = \frac{P_0 e^{0.1t}}{1 + \frac{P_0}{100}(e^{0.1t} - 1)}$$

b) Show that if $P_0 > 0$, then $\lim_{t\to\infty} P(t) = 100$, so that in the long-run, the population described by this logistic equation converges to its carrying capacity $K = 100$.

EXERCISE 5.2 Consider the differential equation $f''(t) = -4f(t)$.

a) Check that the function $f(t) = \cos(2t)$ is a solution.
b) Check that the function $f(t) = 3\sin(2t)$ is a solution.
c) More generally, check that for all constants A, B, the following function is a solution:

$$f(t) = A\cos(2t) + B\sin(2t).$$

EXERCISE 5.3 Find a solution to the differential equation $f''(t) = -4f(t)$ satisfying the initial conditions $f(0) = 1$ and $f'(0) = -1$.

EXERCISE 5.4 Fix a complex number α and consider the differential equation $f'(z) = \alpha f(z)$. Mimic Example 5.1 to show that the unique solution satisfying $f(0) = a_0$ is given by the function $f(z) = a_0 \exp(\alpha z)$.

EXERCISE 5.5 Now consider the equation $f''(z) = -f(z)$. Mimic Example 5.2 to show that the unique solution satisfying $f(0) = a_0$ and

$f'(0) = a_1$ is given by

$$f(z) = a_0 \cos(z) + a_1 \sin(z).$$

Note that taking $a_1 = ia_0$ yields the solution $a_0 \exp(iz)$ to the differential equation $f'(z) = if(z)$.

EXERCISE 5.6 Show that for all constants A_1, A_2, the function $f(z)$ defined below is a solution to $f''(z) + f'(z) + f(z) = 0$:

$$f(z) = A_1 \exp(\alpha_1 z) + A_2 \exp(\alpha_2 z).$$

(Here, α_1 and α_2 are the roots of the polynomial $z^2 + z + 1$.)

EXERCISE 5.7 Solve the following pair of equations for A_1 and A_2:

$$\begin{aligned} A_1 + A_2 &= a_0 \\ A_1\alpha_1 + A_2\alpha_2 &= a_1. \end{aligned}$$

You should find that

$$A_1 = \frac{a_0\alpha_2 - a_1}{\alpha_2 - \alpha_1}, \qquad A_2 = \frac{a_0\alpha_1 - a_1}{\alpha_1 - \alpha_2}.$$

EXERCISE 5.8 Show by direct computation that $x(t) = t\exp(-\frac{rt}{2})$ is a solution to the critically damped oscillator differential equation

$$x''(t) + rx'(t) + \frac{r^2}{4}x(t) = 0.$$

EXERCISE 5.9 Use the series definitions to show that

$$(\sinh(z))' = \cosh(z) \quad \text{and} \quad (\cosh(z))' = \sinh(z).$$

EXERCISE 5.10

a) Using the power series definitions, show that

$$\exp(z) = \cosh(z) + \sinh(z).$$

b) Show that $\cosh(z)$ is an even function and $\sinh(z)$ is an odd function:

$$\cosh(-z) = \cosh(z), \qquad \sinh(-z) = -\sinh(z).$$

c) Use parts a) and b) to show that

$$\cosh(z) = \frac{\exp(z) + \exp(-z)}{2} \qquad \text{and} \qquad \sinh(z) = \frac{\exp(z) - \exp(-z)}{2}.$$

EXERCISE 5.11 Convince yourself that the area of the shaded region on the left side of Figure 5.11 is equal to $\theta/2$. Thus, the parameter θ in $[0, 2\pi)$ for the unit circle may be interpreted as twice the area of the shaded region swept out by the red radius pointing to $(\cos(\theta), \sin(\theta))$.

5.6. Problems

5.1. Consider the population growth differential equation with initial condition:

$$P'(t) = rP(t), \qquad P(0) = P_0.$$

Suppose that $f(t)$ is a solution satisfying the initial condition, so that $f'(t) = rf(t)$ and $f(0) = P_0$. Then consider the function $g(t)$ defined by

$$g(t) = \frac{f(t)}{e^{rt}}.$$

Show that $g'(t) = 0$ for all t, and hence that $g(t)$ is a constant function. Then show that in fact $g(t) = P_0$ for all t, and conclude that

$$f(t) = P_0 e^{rt}.$$

5.2. Consider the differential equation $f'(t) = -3f(t) + e^{-2t}$. Show that for any constant C, the following function is a solution:

$$f(t) = e^{-2t} + Ce^{-3t}.$$

5.3. Consider the differential equation $f'(t) = \sin(f(t))$.

a) Show that for any constant C, the functions $f(t) = \pm 2\,\mathrm{arccot}(e^{C-t})$ are solutions. (Observe that these solutions never change sign: the $+$ yields a positive solution, and the $-$ yields a negative solution.)
b) Show that the constant C is the unique value of t for which $f(t)$ equals $\pm\frac{\pi}{2}$ (and hence the unique value of t for which $f'(t)$ equals ± 1).

5.4. Fix a nonzero complex number α and consider the differential equation $f''(z) = \alpha f(z)$. Mimic Example 5.2 to show that the unique solution satisfying $f(0) = a_0$ and $f'(0) = a_1$ is given by

$$f(z) = a_0 \cosh(\sqrt{\alpha}\, z) + \frac{a_1}{\sqrt{\alpha}} \sinh(\sqrt{\alpha}\, z).$$

Does it matter which square root of α you use?

5.5. Fix complex numbers b and c and consider the differential equation

$$f''(z) + bf'(z) + cf(z) = 0.$$

Under the assumption that $b^2 \neq 4c$, show that there are exactly two values of α that yield a solution $\exp(\alpha z)$ to the differential equation. Adapt the method of Example 5.4 to find the solution with initial conditions $f(0) = 1$ and $f'(0) = 0$.

5.6. This problem concerns Bessel's differential equation of order $k = 1$:

$$z^2 f''(z) + zf'(z) + (z^2 - 1)f(z) = 0.$$

Mimic the method of Example 5.5 to find a power series solution. Your answer should involve one free constant a_1. If you set $a_1 = \frac{1}{2}$ in your solution, you should find the *Bessel function of the first kind of order 1*:

$$J_1(z) = \frac{z}{2}\sum_{n=0}^{\infty}(-1)^n \frac{z^{2n}}{2^{2n} n!(n+1)!}.$$

See Figure 5.13 for the graph of the real version of this function.

5.7. Fix a nonnegative integer k and consider Bessel's differential equation of order k:

$$z^2 f''(z) + zf'(z) + (z^2 - k^2)f(z) = 0.$$

Mimic the method of Example 5.5 to find a power series solution. Your answer should involve one free constant a_k. If you set $a_k = \frac{1}{2^k k!}$ in your solution, you should find the *Bessel function of the first kind of order k*:

$$J_k(z) = \frac{z^k}{2^k}\sum_{n=0}^{\infty}(-1)^n \frac{z^{2n}}{2^{2n} n!(n+k)!}.$$

See Figure 5.13 for the graphs of the real versions of some of these functions.

5.8. Fix a nonnegative integer $k \geq 0$, and consider the differential equation

$$zf''(z) + (1 - z)f'(z) + kf(z) = 0.$$

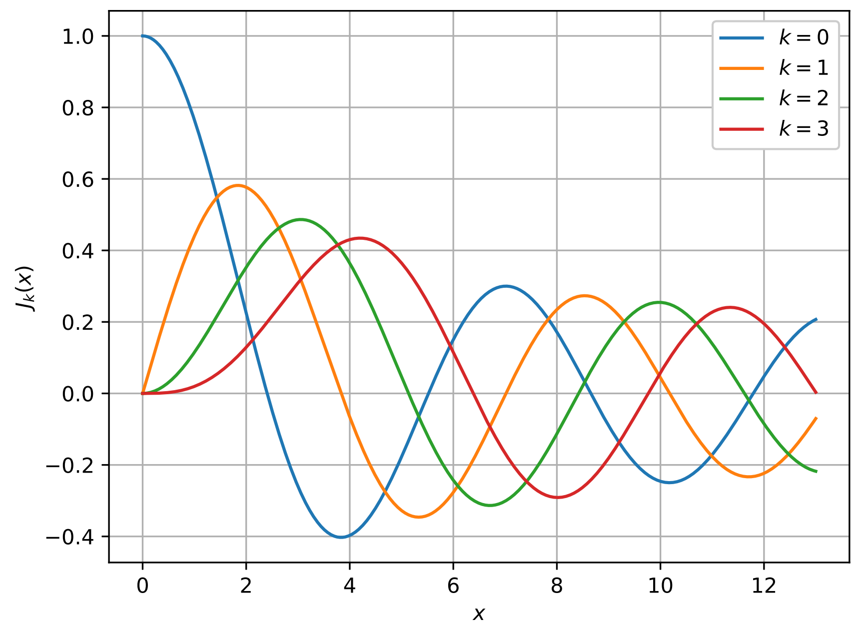


FIGURE 5.13. Graphs of some real Bessel functions $J_k(x)$

Mimic the method of Example 5.5 to find the unique solution $f(z)$ with $f(0) = 1$. Explain why your answer is actually a polynomial of degree k. It is called the *kth Laguerre polynomial*.