Amazing and Aesthetic Aspects of Analysis:

On the incredible infinite

(A Course in Undergraduate Analysis, Fall 2006)

$$
\begin{aligned}
\frac{\pi^2}{6} &= \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \\
&= \frac{2^2}{2^2-1} \cdot \frac{3^2}{3^2-1} \cdot \frac{5^2}{5^2-1} \cdot \frac{7^2}{7^2-1} \cdot \frac{11^2}{11^2-1} \cdots \\
&= \cfrac{1}{0^2 + 1^2 - \cfrac{1^4}{1^2 + 2^2 - \cfrac{2^4}{2^2 + 3^2 - \cfrac{3^4}{3^2 + 4^2 - \cfrac{4^4}{4^2 + 5^2 - \ddots}}}}}
\end{aligned}
$$

Paul Loya

Contents

Preface i

Acknowledgement iii

Some of the most beautiful formulæ in the world v

A word to the student vii

Part 1. Some standard curriculum 1

Chapter 1. Sets, functions, and proofs 3
1.1. The algebra of sets and the language of mathematics 4
1.2. Set theory and mathematical statements 11
1.3. What are functions? 15

Chapter 2. Numbers, numbers, and more numbers 21
2.1. The natural numbers 22
2.2. The principle of mathematical induction 27
2.3. The integers 35
2.4. Primes and the fundamental theorem of arithmetic 41
2.5. Decimal representations of integers 49
2.6. Real numbers: Rational and "mostly" irrational 53
2.7. The completeness axiom of $\mathbb{R}$ and its consequences 63
2.8. m-dimensional Euclidean space 72
2.9. The complex number system 79
2.10. Cardinality and "most" real numbers are transcendental 83

Chapter 3. Infinite sequences of real and complex numbers 93
3.1. Convergence and ε-N arguments for limits of sequences 94
3.2. A potpourri of limit properties for sequences 102
3.3. The monotone criteria, the Bolzano-Weierstrass theorem, and e 111
3.4. Completeness and the Cauchy criteria for convergence 117
3.5. Baby infinite series 123
3.6. Absolute convergence and a potpourri of convergence tests 131
3.7. Tannery's theorem, the exponential function, and the number e 138
3.8. Decimals and "most" numbers are transcendental á la Cantor 146

Chapter 4. Limits, continuity, and elementary functions 153
4.1. Convergence and ε-δ arguments for limits of functions 154
4.2. A potpourri of limit properties for functions 160
4.3. Continuity, Thomae's function, and Volterra's theorem 166

4.4. Compactness, connectedness, and continuous functions 172
4.5. Monotone functions and their inverses 182
4.6. Exponentials, logs, Euler and Mascheroni, and the ζ-function 187
4.7. The trig functions, the number π, and which is larger, π^e or e^π? 198
4.8. ★ Three proofs of the fundamental theorem of algebra (FTA) 210
4.9. The inverse trigonometric functions and the complex logarithm 217
4.10. ★ The amazing π and its computations from ancient times 226

Chapter 5. Some of the most beautiful formulæ in the world 237
5.1. ★ Euler, Wallis, and Viète 240
5.2. ★ Euler, Gregory, Leibniz, and Madhava 249
5.3. ★ Euler's formula for $\zeta(2k)$ 258

Part 2. Extracurricular activities 267

Chapter 6. Advanced theory of infinite series 269
6.1. Summation by parts, bounded variation, and alternating series 271
6.2. Liminfs/sups, ratio/roots, and power series 279
6.3. A potpourri of ratio-type tests and "big $\mathcal{O}$" notation 290
6.4. Some pretty powerful properties of power series 295
6.5. Double sequences, double series, and a ζ-function identity 300
6.6. Rearrangements and multiplication of power series 311
6.7. ★ Proofs that $\sum 1/p$ diverges 320
6.8. Composition of power series and Bernoulli and Euler numbers 325
6.9. The logarithmic, binomial, arctangent series, and γ 332
6.10. ★ π, Euler, Fibonacci, Leibniz, Madhava, and Machin 340
6.11. ★ Another proof that $\pi^2/6 = \sum_{n=1}^\infty 1/n^2$ (The Basel problem) 344

Chapter 7. More on the infinite: Products and partial fractions 349
7.1. Introduction to infinite products 350
7.2. Absolute convergence for infinite products 355
7.3. Euler, Tannery, and Wallis: Product expansions galore 359
7.4. Partial fraction expansions of the trigonometric functions 366
7.5. ★ More proofs that $\pi^2/6 = \sum_{n=1}^\infty 1/n^2$ 370
7.6. ★ Riemann's remarkable ζ-function, probability, and $\pi^2/6$ 373
7.7. ★ Some of the most beautiful formulæ in the world IV 382

Chapter 8. Infinite continued fractions 389
8.1. Introduction to continued fractions 390
8.2. ★ Some of the most beautiful formulæ in the world V 394
8.3. Recurrence relations, Diophantus' tomb, and shipwrecked sailors 403
8.4. Convergence theorems for infinite continued fractions 411
8.5. Diophantine approximations and the mystery of π solved! 422
8.6. ★ Continued fractions and calendars, and math and music 433
8.7. The elementary functions and the irrationality of $e^{p/q}$ 437
8.8. Quadratic irrationals and periodic continued fractions 446
8.9. Archimedes' crazy cattle conundrum and diophantine equations 456
8.10. Epilogue: Transcendental numbers, π, e, and where's calculus? 464

Bibliography 471

Index 481

Preface

I have truly enjoyed writing this book. Admittedly, some of the writing is too overdone (e.g. overdoing alliteration at times), but what can I say, I was having fun. The "starred" sections of the book are meant to be "just for fun" and don't interfere with other sections (besides perhaps other starred sections). Most of the quotes that you'll find in these pages are taken from the website `http://www-gap.dcs.st-and.ac.uk/~history/Quotations/`

This is a first draft, so *please* email me any errors, suggestions, comments etc. about the book to

`paul@math.binghamton.edu`.

The overarching goals of this textbook are similar to any advanced math textbook, regardless of the subject:

GOALS OF THIS TEXTBOOK. THE STUDENT WILL BE ABLE TO ...

- comprehend and write mathematical reasonings and proofs.
- wield the language of mathematics in a precise and effective manner.
- state the fundamental ideas, axioms, definitions, and theorems upon which real analysis is built and flourishes.
- articulate the need for abstraction and the development of mathematical tools and techniques in a general setting.

The objectives of this book make up the framework of how these goals will be accomplished, and more or less follow the chapter headings:

OBJECTIVES OF THIS TEXTBOOK. THE STUDENT WILL BE ABLE TO ...

- identify the interconnections between set theory and mathematical statements and proofs.
- state the fundamental axioms of the natural, integer, and real number systems and how the completeness axiom of the real number system distinguishes this system from the rational system in a powerful way.
- apply the rigorous ε-N definition of convergence for sequences and series and recognize monotone and Cauchy sequences.
- apply the rigorous ε-δ definition of limits for functions and continuity and the fundamental theorems of continuous functions.
- determine the convergence and properties of an infinite series, product, or continued fraction using various tests.
- identify series, product, and continued fraction formulæ for the various elementary functions and constants.

I'd like to thank Brett Bernstein for looking over the notes and gave many valuable suggestions.

Finally, some last words about my book. This not a history book (but we try to talk history throughout this book) and this not a "little" book like Herbert Westren Turnbull's book *The Great Mathematicians*, but like Turnbull, I do hope

> *If this little book perhaps may bring to some, whose acquaintance with mathematics is full of toil and drudgery, a knowledge of those great spirits who have found in it an inspiration and delight, the story has not been told in vain. There is a largeness about mathematics that transcends race and time: mathematics may humbly help in the market-place, but it also reaches to the stars. To one, mathematics is a game (but what a game!) and to another it is the handmaiden of theology. The greatest mathematics has the simplicity and inevitableness of supreme poetry and music, standing on the borderland of all that is wonderful in Science, and all that is beautiful in Art. Mathematics transfigures the fortuitous concourse of atoms into the tracery of the finger of God.*
>
> *Herbert Westren Turnbull (1885–1961). Quoted from* [**225**, p. 141].

Paul Loya
Binghamton University, Vestal Parkway, Binghamton, NY 13902
paul@math.binghamton.edu

Soli Deo Gloria

Acknowledgement

To Jesus, my Lord, Savior and Friend.

Some of the most beautiful formulæ in the world

Here is a very small sample of the many beautiful formulas we'll prove in this book involving some of the main characters we'll meet in our journey.

$$e = 2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \cfrac{5}{5 + \ddots}}}} \qquad \text{(Euler; §5.2)}$$

$$\gamma = \sum_{n=2}^{\infty} \frac{(-1)^n}{n} \zeta(n) \qquad \text{(Euler; §6.9)}$$

$$\log 2 = \frac{2}{1+\sqrt{2}} \cdot \frac{2}{1+\sqrt{\sqrt{2}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{2}}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{\sqrt{2}}}}} \cdots \qquad \text{(Seidel; §7.1)}$$

$$\Phi = \frac{1+\sqrt{5}}{2} = \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}}}} \qquad \text{(§3.3)}$$

$$\Phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}} \qquad \text{(§3.4)}$$

$$\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \qquad \text{(Euler; §6.11)}$$

$$\frac{\pi^4}{90} = \frac{1}{1^4} + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots \qquad \text{(Euler; §7.5)}$$

$$\frac{\pi}{4} = \frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots \qquad \text{(Gregory-Leibniz-Madhava; §6.10)}$$

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots \qquad \text{(Viète; §4.10)}$$

$$\frac{\pi}{2} = \frac{1}{1} \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{7} \cdots \qquad \text{(Wallis; §6.10)}$$

$$\frac{4}{\pi} = 1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ddots}}}} \qquad \text{(Lord Brouncker; §5.2)}$$

$$e^x = \cfrac{1}{1 - \cfrac{2x}{x + 2 + \cfrac{x^2}{6 + \cfrac{x^2}{10 + \cfrac{x^2}{14 + \ddots}}}}} \qquad \text{(Euler; §8.7)}$$

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) \qquad \text{(Euler; §7.3)}$$

$$\zeta(z) = \prod \left(1 - \frac{1}{p^z}\right)^{-1} = \prod \frac{p^z}{p^z - 1} \qquad \text{(Euler–Riemann; §7.6)}$$

A word to the student

One can imagine mathematics as a movie with exciting scenes, action, plots, etc. ... There are a couple things you can do. First, you can simply sit back and watch the movie playing out. Second, you can take an active role in shaping the movie. A mathematician does both at times, but is more the actor rather than the observer. I recommend you be the actor in the great mathematics movie. To do so, I recommend you read this book with a pencil and paper at hand writing down definitions, working through examples filling in any missing details, and of course doing exercises (even the ones that are not assigned).[1] Of course, please feel free to mark up the book as much as you wish with remarks and highlighting and even corrections if you find a typo or error. (Just let me know if you find one!)

[1] There are many footnotes in this book. Most are quotes from famous mathematicians and others are remarks that I might say to you if I were reading the book with you. All footnotes may be ignored if you wish!

Part 1

Some standard curriculum

CHAPTER 1

Sets, functions, and proofs

Mathematics is not a deductive science — that's a cliche. When you try to prove a theorem, you don't just list the hypotheses, and then start to reason. What you do is trial and error, experimentation, guesswork.
Paul R. Halmos (1916–), I want to be a Mathematician [**92**].

In this chapter we start being "mathematicians", that is, we start doing proofs. One of the goals of this text is to get you proving mathematical statements in real analysis. Set theory provides a safe environment in which to learn about math statements, "if ... then", "if and only if", etc., and to learn the logic behind proofs. Actually, I assume that most of you have had some exposure to sets, so many proofs in this chapter are "left to the reader"; the real meat comes in the next chapter.

The students at Binghamton University, the people in your family, your pets, the food in your refrigerator are all examples of sets of objects. Mathematically, a set is defined by some property or attribute that an object must have or must not have; if an object has the property, then it's in the set. For example, the collection of all registered students at Binghamton University who are signed up for real analysis forms a set. (A BU student is either signed up for real analysis or not.) For an example of a property that cannot be used to define a set, try to answer the following question proposed by Bertrand Russell (1872–1970) in 1918:

A puzzle for the student: *A barber in a local town puts up a sign saying that he shaves only those people who do not shave themselves. "Who, then, shaves the barber?"*

Try to answer this question. (Does the barber shave himself or does someone else shave him?) Any case, the idea of a set is perhaps the most fundamental idea in all of mathematics. Sets can be combined to form other sets and the study of such operators is called the *algebra of sets*, which we cover in Section 1.1. In Section 1.2 we look at the relationship between set theory and the language of mathematics. Second to sets in fundamental importance is the idea of a function, which we cover in Section 1.3. In order to illustrate relevant examples of sets, we shall presume elementary knowledge of the real numbers. A thorough discussion of real numbers is left for the next chapter.

This chapter is short on purpose since we do not want to spend too much time on set theory so as to start real analysis ASAP. In the words of Paul Halmos [**91**, p. vi], "... general set theory is pretty trivial stuff really, but, if you want to be a mathematician, you need some, and here it is; read it, absorb it, and forget it."

Chapter 1 objectives: The student will be able to ...

- manipulate and create new sets from old ones using the algebra of sets
- identify the interconnections between set theory and math statements/proofs.
- Define functions and the operations of functions on sets.

1.1. The algebra of sets and the language of mathematics

In this section we study sets and various operations, referred to as the *algebra of sets*, to form other sets. We shall see that the algebra of sets is indispensable in many branches of mathematics such as the study of topology in later chapters. Set theory also provides the language by which mathematics and logic are built.

1.1.1. Sets and intervals. A **set** is a collection of definite, well-distinguished objects, also called elements, which are usually defined by a conditional statement or simply by listing their elements. All sets and objects that we deal with have the property that given an object and a set, there must be a definite "yes" or "no" answer to whether or not the object is in the set, because otherwise paradoxes can arise as seen in the barber paradox; see also Problem 4 for another puzzle.

Example 1.1. Sets where we can list the elements include

$$\mathbb{N} := \{1, 2, 3, 4, 5, 6, 7, \ldots\} \quad \text{and} \quad \mathbb{Z} := \{\cdots, -2, -1, 0, 1, 2, \cdots\},$$

the **natural numbers** and **integers**, respectively. Here, the symbol ":=" means that the symbol on the left is by definition the expression on the right and we usually read ":=" as "equals by definition".[1]

Example 1.2. We can define the **rational numbers** by the conditional statement

$$\mathbb{Q} := \left\{x \in \mathbb{R}\,;\, x = \frac{a}{b}, \ \ \text{where } a, b \in \mathbb{Z} \text{ and } b \neq 0\right\}.$$

Here, $\mathbb{R}$ denotes the set of real numbers and the semicolon should be read "such that". So, $\mathbb{Q}$ is the set of all real numbers x such that x can be written as a ratio $x = a/b$ where a and b are integers with b not zero.

Example 1.3. The **empty set** is a set with no elements — think of an empty clear plastic bag. We denote this empty set by $\varnothing$. (In the next subsection we prove that there is only one empty set.)

Example 1.4. Intervals provide many examples of sets defined by conditional statements. Let a and b be real numbers with $a \le b$. Then the set

$$\{x \in \mathbb{R}\,;\, a < x < b\}$$

is called an **open interval** and is often denoted by (a, b). If $a = b$, then there are no real numbers between a and b, so $(a, a) = \varnothing$. The set

$$\{x \in \mathbb{R}\,;\, a \le x \le b\}$$

is called a **closed interval** and is denoted by $[a, b]$. There are also half open and closed intervals,

$$\{x \in \mathbb{R}\,;\, a < x \le b\}, \qquad \{x \in \mathbb{R}\,;\, a \le x < b\},$$

called **left-half open** and **right-half open** intervals and are denoted by $(a, b]$ and $[a, b)$, respectively. The points a and b are called the **end points** of the intervals. There are also infinite intervals. The sets

$$\{x \in \mathbb{R}\,;\, x < a\}, \qquad \{x \in \mathbb{R}\,;\, a < x\}$$

[1] *The errors of definitions multiply themselves according as the reckoning proceeds; and lead men into absurdities, which at last they see but cannot avoid, without reckoning anew from the beginning. Thomas Hobbes (1588–1679)* [**160**].

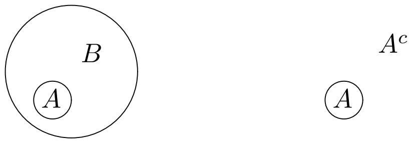


FIGURE 1.1. The left-hand side displays a subset. The right-hand side deals with complements that we'll look at in Section 1.1.3.

are open intervals, denoted by $(-\infty, a)$ and (a, ∞), respectively, and

$$\{x \in \mathbb{R}\,;\, x \leq a\}, \qquad \{x \in \mathbb{R}\,;\, a \leq x\}$$

are closed intervals, denoted by $(-\infty, a]$ and $[a, \infty)$, respectively. Note that the sideways eight symbol ∞ for "infinity," introduced in 1655 by John Wallis (1616–1703) [**45**, p. 44], is just that, a symbol, and is not to be taken to be a real number. The real line is itself an interval, namely $\mathbb{R} = (-\infty, \infty)$.

1.1.2. Subsets and "if ... then" statements. If a is belongs to a set A, then we usually say a is in A and we write $a \in A$ or if a does not belong to A, then we write $a \notin A$. If each element of a set A is also an element of a set B, we write $A \subseteq B$ and say that A is a **subset** of, or contained in, B. Thus, $A \subseteq B$ means if a is in A, then also a is in B. See Figure 1.1. If A not a subset of B, we write $A \not\subseteq B$.

Example 1.5. $\mathbb{N} \subseteq \mathbb{Z}$ since every natural number is also an integer and $\mathbb{Z} \subseteq \mathbb{R}$ since every integer is also a real number, but $\mathbb{R} \not\subseteq \mathbb{Z}$ because not every real number is an integer.

To say that two sets A and B are the same just means that they contain exactly the same elements; in other words, every element in A is also in B (that is, $A \subseteq B$) and also every element in B is also in A (that is, $B \subseteq A$). Thus, we define

$$\boxed{A = B \quad \text{means that } A \subseteq B \text{ and } B \subseteq A.}$$

A set that is a subset of every set is the empty set $\varnothing$. To see that $\varnothing$ is a subset of every set, let A be a set. We must show that the statement if $x \in \varnothing$, then $x \in A$ is true. However, the part "$x \in \varnothing$" of this statement sounds strange because the empty set has nothing in it, so $x \in \varnothing$ is an untrue statement to begin with. Before evaluating the statement "If $x \in \varnothing$, then $x \in A$," let us first discuss general "If ... then" statements. Consider the following following statement made by Joe:

If Professor Loya cancels class on Friday, then I'm driving to New York City.

Obviously, Joe told the truth if indeed I cancelled class and he headed off to NYC. What if I did not cancel class but Joe still went to NYC, did Joe tell the truth, lie, or neither: his statement simply does not apply? Mathematicians would say that Joe told the truth (regardless of the outcome, Joe went to NYC or stayed in Binghamton). He only said *if* the professor cancels class, *then* he would drive to NYC. All bets are off if the professor does not cancel class. This is the standing *convention* mathematicians take for any "If ... then" statement. Thus, given statements P and Q, we consider the statement "If P, then Q" to be true if the statement P is true, then the statement Q is also true, and we also regard it as being true if the statement P is false whether or not the statement Q is true or false. There is no such thing as a "neither statement" in this book.[2]

Now back to our problem. We want to prove that if $x \in \varnothing$, then $x \in A$. Since $x \in \varnothing$ is untrue, by our *convention*, the statement "if $x \in \varnothing$, then $x \in A$" is true by default. Thus, $\varnothing \subseteq A$. We can also see that there is only one empty set, for suppose that $\varnothing'$ is another empty set. Then the same (silly) argument that we just did for $\varnothing$ shows that $\varnothing'$ is also a subset of every set. Now to say that $\varnothing = \varnothing'$, we must show that $\varnothing \subseteq \varnothing'$ and $\varnothing' \subseteq \varnothing$. But $\varnothing \subseteq \varnothing'$ holds because $\varnothing$ is a subset of every set and $\varnothing' \subseteq \varnothing$ holds because $\varnothing'$ is a subset of every set. Therefore, $\varnothing = \varnothing'$.

There is another, perhaps easier, way to see that $\varnothing$ is a subset of any set by invoking the "contrapositive". Consider again the statement that $A \subseteq B$:

$$\text{(1)} \quad \text{If } x \in A, \text{ then } x \in B.$$

This is equivalent to the **contrapositive** statement

$$\text{(2)} \quad \text{If } x \notin B, \text{ then } x \notin A.$$

Indeed, suppose that statement (1) holds, that is, $A \subseteq B$. We shall prove that statement (2) holds. So, let us assume that $x \notin B$ is true; is true that $x \notin A$?[3] Well, the object x is either in A or it's not. If $x \in A$, then, since $A \subseteq B$, we must have $x \in B$. However, we know that $x \notin B$, and so $x \in A$ is not the valid option, and therefore $x \notin A$. Assume now that statement (2) holds: If $x \notin B$, then $x \notin A$. We shall prove that statement (1) holds, that is, $A \subseteq B$. So, let $x \in A$. We must prove that $x \in B$. Well, either $x \in B$ or it's not. If $x \notin B$, then we know that $x \notin A$. However, we are given that $x \in A$, so $x \notin B$ is not the correct option, therefore, the other option $x \in B$ must be true. Therefore, (1) and (2) really say the same thing. We now prove that $\varnothing \subseteq A$ for any given set A. Assume that $x \notin A$. According to (2), we must prove that $x \notin \varnothing$. But this last statement is true because $\varnothing$ does not contain anything, so $x \notin \varnothing$ is certainly true. Thus, $\varnothing \subseteq A$.

The following theorem states an important law of sets.

THEOREM 1.1 (**Transitive law**). *If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.*

PROOF. Suppose that $A \subseteq B$ and $B \subseteq C$. We need to prove that $A \subseteq C$, which by definition means that if $x \in A$, then $x \in C$. So, let x be in A; we need to show that x is also in C. Since x is in A and $A \subseteq B$, we know that x is also in B. Now $B \subseteq C$, and therefore x is also in C. In conclusion, we have proved that if $x \in A$, then $x \in C$, which is exactly what we wanted to prove. □

Finally, we remark that the **power set** of a given set A is the collection consisting of all subsets of A, which we usually denote by $\mathscr{P}(A)$.

Example 1.6.

$$\mathscr{P}(\{e, \pi\}) = \big\{\varnothing, \{e\}, \{\pi\}, \{e, \pi\}\big\}.$$

[2] Later in your math career you will find some "neither statements" such as e.g. the continuum hypothesis . . . but this is another story!

[3] Recall our *convention* that for a false statement P, we always consider a statement "If P, then Q" to be true regardless of the validity of the statement Q. Therefore, "$x \notin B$" is false automatically makes the statement (2) true regardless of the validity of the statement "$x \notin A$", so in order to prove statement (2) is true, we might as well assume that the statement "$x \notin B$" is true and try to show that the statement "$x \notin A$" is also true.

1.1.3. Unions, "or" statements, intersections, and set differences. Given two sets A and B, their **union**, denoted $A \cup B$, is the set of elements that are in A or B:

$$\boxed{A \cup B := \{x\,;\, x \in A \ \text{ or } \ x \in B\}.}$$

Here we come to another difference between English and mathematical language. Let's say that your parents come to visit you on campus and your dad asks you:

Would you like to go to McDonald's or Burger King?

By "or," your dad means that you can choose only one of the two choices, but not both. At the restaurant, your mom asks:

Would you like to have ketchup or mustard?

Now by "or" in this case, your mom means that you can choose ketchup, mustard, or both if you want. Mathematicians always follow mom's meaning of "or" (mom is always right ☺)! Thus, $A \cup B$, is the set of elements that are in A or B, where "or" means in A, in B, or in both A and B.

Example 1.7.

$$\{0, 1, e, i\} \cup \{e, i, \pi, \sqrt{2}\} = \{0, 1, e, i, \pi, \sqrt{2}\}.$$

The **intersection** of two sets A and B, denoted by $A \cap B$, is the set of elements that are in both A and B:

$$\boxed{A \cap B := \{x\,;\, x \in A \ \text{ and } \ x \in B\}.}$$

(Here, "and" means just what you think it means.)

Example 1.8.

$$\{0, 1, e, i\} \cap \{e, i, \pi, \sqrt{2}\} = \{e, i\}.$$

If the sets A and B have no elements in common, then $A \cap B = \varnothing$, and the sets are said to be **disjoint**. Here are some properties of unions and intersections, the proofs of which we leave mostly to the reader.

THEOREM 1.2. *Unions and intersections are commutative and associative in the sense that if A, B, and C are sets, then*
(1) $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$.
(2) $(A \cup B) \cup C = A \cup (B \cup C)$ *and* $(A \cap B) \cap C = A \cap (B \cap C)$.

PROOF. Consider the proof that $A \cup B = B \cup A$. By definition of equality of sets, we must show that $A \cup B \subseteq B \cup A$ and $B \cup A \subseteq A \cup B$. To prove that $A \cup B \subseteq B \cup A$, let x be in $A \cup B$. Then by definition of union, $x \in A$ or $x \in B$. This of course is the same thing as $x \in B$ or $x \in A$. Therefore, x is in $B \cup A$. The proof that $B \cup A \subseteq A \cup B$ is similar. Therefore, $A \cup B = B \cup A$. We leave the proof that $A \cap B = B \cap A$ to the reader. We also leave the proof of *(2)* to the reader. □

Our last operation on sets is the **set difference** $A \setminus B$ (read "A take away B" or the "complement of B in A"), which is the set of elements of A that do not belong to B. Thus,

$$\boxed{A \setminus B := \{x\,;\, x \in A \text{ and } x \notin B\}.}$$

Example 1.9.

$$\{0, 1, e, i\} \setminus \{e, i, \pi, \sqrt{2}\} = \{0, 1\}.$$

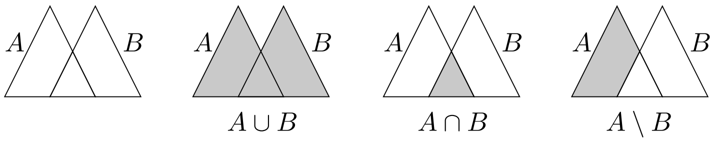


FIGURE 1.2. Visualization of the various set operations. Here, A and B are overlapping triangles.

It is *always* assumed that in any given situation we working with subsets of some underlying "universal" set X. Given any subset A of X, we denote $X \setminus A$, the set of elements in X that are outside of A, by A^c, called the **complement** of A; see Figure 1.1. Therefore,

$$\boxed{A^c := X \setminus A = \{x \in X \,;\, x \notin A\}.}$$

Example 1.10. Let us take our "universe" to be $\mathbb{R}$. Then,

$$(-\infty, 1]^c = \{x \in \mathbb{R}\,;\, x \notin (-\infty, 1]\} = (1, \infty),$$

and

$$[0, 1]^c = \{x \in \mathbb{R}\,;\, x \notin [0, 1]\} = (-\infty, 0) \cup (1, \infty).$$

In any given situation, the universal set X will always be clear from context, either because it is stated what X is, or because we are working in, say a section dealing with real numbers only, so $\mathbb{R}$ is by default the universal set. Otherwise, we assume that X is just "there" but simply not stated. For pictorial representations of union, intersection, and set difference, see Figure 1.2; these pictures are called **Venn diagrams** after John Venn (1834–1923) who introduced them.

1.1.4. Arbitrary unions and intersections. We can also consider arbitrary (finite or infinite) unions and intersections. Let I be a nonempty set and assume that for each $\alpha \in I$, there corresponds a set A_α. The sets A_α where $\alpha \in I$ are said to be a **family** of sets **indexed** by I, which we often denote by $\{A_\alpha \,;\, \alpha \in I\}$. An index set that shows up quite often is $I = \mathbb{N}$; in this case we usually call $\{A_n \,;\, n \in \mathbb{N}\}$ a **sequence** of sets.

Example 1.11. For example, $A_1 := [0, 1]$, $A_2 := [0, 1/2]$, $A_3 := [0, 1/3]$, and in general,

$$A_n := \left[0, \frac{1}{n}\right] = \left\{x \in \mathbb{R}\,;\, 0 \leq x \leq \frac{1}{n}\right\},$$

form a family of sets indexed by $\mathbb{N}$ (or a sequence of sets). See Figure 1.3 for a picture of these sets.

How do we define the union of all the sets A_α in a family $\{A_\alpha \,;\, \alpha \in I\}$? Consider the case of two sets A and B. We can write

$$\begin{aligned} A \cup B &= \{x \,;\, x \in A \quad \text{or} \quad x \in B\} \\ &= \{x \,;\, x \text{ is in at least one of the sets on the left-hand side}\}. \end{aligned}$$

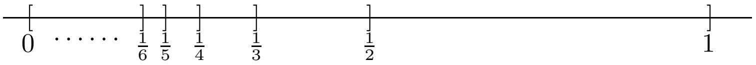


FIGURE 1.3. The sequence of sets $A_n = [0, 1/n]$ for $n \in \mathbb{N}$.

With this as motivation, we define the union of all the sets A_α to be

$$\boxed{\bigcup_{\alpha \in I} A_\alpha := \{x\,;\, x \in A_\alpha \text{ for at least one } \alpha \in I\}.}$$

To simplify notation, we sometimes just write $\bigcup A_\alpha$ or $\bigcup_\alpha A_\alpha$ for the left-hand side.

Example 1.12. For the sequence $\{A_n\,;\, n \in \mathbb{N}\}$ where $A_n = [0, 1/n]$, by staring at Figure 1.3 we see that

$$\bigcup_{n \in \mathbb{N}} A_n := \{x\,;\, x \in [0, 1/n] \text{ for at least one } n \in \mathbb{N}\} = [0, 1].$$

We how do we define the intersection of all the sets A_α in a family $\{A_\alpha\,;\, \alpha \in I\}$? Consider the case of two sets A and B. We can write

$$\begin{aligned} A \cap B &= \{x\,;\, x \in A \quad \text{and} \quad x \in B\} \\ &= \{x\,;\, x \text{ is in every set on the left-hand side}\}. \end{aligned}$$

With this as motivation, we define the intersection of all the sets A_α to be

$$\boxed{\bigcap_{\alpha \in I} A_\alpha := \{x\,;\, x \in A_\alpha \text{ for every } \alpha \in I\}.}$$

Example 1.13. For the sequence $A_n = [0, 1/n]$ in Figure 1.3, we have

$$\bigcap_{n \in \mathbb{N}} A_n := \{x\,;\, x \in [0, 1/n] \text{ for every } n \in \mathbb{N}\} = \{0\}.$$

To simplify notation, we sometimes just write $\bigcap A_\alpha$ or $\bigcap_\alpha A_\alpha$ for the left-hand side. If $I = \{1, 2, \ldots, N\}$ is a finite set of natural numbers, then we usually denote $\bigcup_\alpha A_\alpha$ and $\bigcap_\alpha A_\alpha$ by $\bigcup_{n=1}^N A_n$ and $\bigcap_{n=1}^N A_n$, respectively. If $I = \mathbb{N}$, then we usually denote $\bigcup_\alpha A_\alpha$ and $\bigcap_\alpha A_\alpha$ by $\bigcup_{n=1}^\infty A_n$ and $\bigcap_{n=1}^\infty A_n$, respectively.

Theorem 1.3. *Let A be a set and $\{A_\alpha\}$ be a family of sets. Then union and intersections distribute in the sense that*

$$A \cap \bigcup_\alpha A_\alpha = \bigcup_\alpha (A \cap A_\alpha), \qquad A \cup \bigcap_\alpha A_\alpha = \bigcap_\alpha (A \cup A_\alpha)$$

*and satisfy the Augustus De Morgan (*1806–1871*) laws:*

$$A \setminus \bigcup_\alpha A_\alpha = \bigcap_\alpha (A \setminus A_\alpha), \qquad A \setminus \bigcap_\alpha A_\alpha = \bigcup_\alpha (A \setminus A_\alpha).$$

Proof. We shall leave the first distributive law to the reader and prove the second one. We need to show that $A \cap \bigcup_\alpha A_\alpha = \bigcup_\alpha (A \cap A_\alpha)$, which means that

$$A \cap \bigcup_\alpha A_\alpha \subseteq \bigcup_\alpha (A \cap A_\alpha) \quad \text{and} \quad \bigcup_\alpha (A \cap A_\alpha) \subseteq A \cap \bigcup_\alpha A_\alpha. \tag{1.1}$$

To prove the first inclusion, let $x \in A \cap \bigcup_\alpha A_\alpha$; we must show that $x \in \bigcup_\alpha (A \cap A_\alpha)$. The statement $x \in A \cap \bigcup_\alpha A_\alpha$ means that $x \in A$ and $x \in \bigcup_\alpha A_\alpha$, which means, by the definition of union, $x \in A$ and $x \in A_\alpha$ for some α. Hence, $x \in A \cap A_\alpha$ for some α, which is to say, $x \in \bigcup_\alpha (A \cap A_\alpha)$. Consider now the second inclusion in (1.1). To prove this, let $x \in \bigcup_\alpha (A \cap A_\alpha)$. This means that $x \in A \cap A_\alpha$ for some α. Therefore, by definition of intersection, $x \in A$ and $x \in A_\alpha$ for some α. This means that $x \in A$ and $x \in \bigcup_\alpha A_\alpha$, which is to say, $x \in A \cap \bigcup_\alpha A_\alpha$. In summary, we have established both inclusions in (1.1), which proves the equality of the sets.

We shall prove the first De Morgan law and leave the second to the reader. We need to show that $A \setminus \bigcup_\alpha A_\alpha = \bigcap_\alpha (A \setminus A_\alpha)$, which means that

$$A \setminus \bigcup_\alpha A_\alpha \subseteq \bigcap_\alpha (A \setminus A_\alpha) \quad \text{and} \quad \bigcap_\alpha (A \setminus A_\alpha) \subseteq A \setminus \bigcup_\alpha A_\alpha. \tag{1.2}$$

To prove the first inclusion, let $x \in A \setminus \bigcup_\alpha A_\alpha$. This means $x \in A$ and $x \notin \bigcup_\alpha A_\alpha$. For x not to be in the union, it must be that $x \notin A_\alpha$ for any α whatsoever (because if x happened to be in some A_α, then x would be in the union $\bigcup_\alpha A_\alpha$ which we know x is not). Hence, $x \in A$ and $x \notin A_\alpha$ for all α, in other words, $x \in A \setminus A_\alpha$ for all α, which means that $x \in \bigcap_\alpha (A \setminus A_\alpha)$. We now prove the second inclusion in (1.2). So, let $x \in \bigcap_\alpha (A \setminus A_\alpha)$. This means that $x \in A \setminus A_\alpha$ for all α. Therefore, $x \in A$ and $x \notin A_\alpha$ for all α. Since x is not in any A_α, it follows that $x \notin \bigcup_\alpha A_\alpha$. Therefore, $x \in A$ and $x \notin \bigcup_\alpha A_\alpha$ and hence, $x \in A \setminus \bigcup_\alpha A_\alpha$. In summary, we have established both inclusions in (1.2), which proves the equality of the sets. □

The best way to remember De Morgan's laws is the English versions: The complement of a union is the intersection of the complements and the complement of an intersection is the union of the complements. For a family $\{A_\alpha\}$ consisting of just two sets B and C, the distributive and De Morgan laws are just

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

and

$$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C), \qquad A \setminus (B \cap C) = (A \setminus C) \cup (A \setminus C).$$

Here are some exercises where we ask you to prove statements concerning sets. In Problem 3 it is very helpful to draw Venn diagrams to "see" why the statement should be true. Some advice that is useful throughout this whole book: If you can't see how to prove something after some effort, take a break and come back to the problem later.[4]

EXERCISES 1.1.

1. Prove that $\varnothing = \{x\,;\, x \neq x\}$. True, false, or neither: If $x \in \varnothing$, then real analysis is everyone's favorite class.
2. Prove that for any set A, we have $A \cup \varnothing = A$ and $A \cap \varnothing = \varnothing$.
3. Prove the following statements:
   (a) $A \setminus B = A \cap B^c$.
   (b) $A \cap B = A \setminus (A \setminus B)$.
   (c) $B \cap (A \setminus B) = \varnothing$.
   (d) If $A \subseteq B$, then $B = A \cup (B \setminus A)$.
   (e) $A \cup B = A \cup (B \setminus A)$.
   (f) $A \subseteq A \cup B$ and $A \cap B \subseteq A$.
   (g) If $A \cap B = A \cap C$ and $A \cup C = A \cup B$, then $B = C$.
   (h) $(A \setminus B) \setminus C = (A \setminus C) \setminus (B \setminus C)$

[4] *Finally, two days ago, I succeeded - not on account of my hard efforts, but by the grace of the Lord. Like a sudden flash of lightning, the riddle was solved. I am unable to say what was the conducting thread that connected what I previously knew with what made my success possible. Carl Friedrich Gauss (1777–1855)* [**67**].

4. (**Russell's paradox**)[5] Define a "thing" to be any collection of items. The reason that we use the word "thing" is that these things are not sets. Let
$$\mathscr{B} = \{\text{ "things" } A\,;\, A \notin A\},$$
that is, $\mathscr{B}$ is the collection of all "things" that do not contain themselves. Questions: Is $\mathscr{B}$ a "thing"? Is $\mathscr{B} \in \mathscr{B}$ or is $\mathscr{B} \notin \mathscr{B}$? Is $\mathscr{B}$ a set?
5. Find
$$(a) \bigcup_{n=1}^{\infty} \left(0, \frac{1}{n}\right), \quad (b) \bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right), \quad (c) \bigcup_{n=1}^{\infty} \left(\frac{1}{2^n}, \frac{1}{2^{n-1}}\right),$$
$$(d) \bigcap_{n=1}^{\infty} \left(\frac{1}{2^n}, \frac{1}{2^{n-1}}\right), \quad (e) \bigcap_{\alpha \in \mathbb{R}} (\alpha, \infty), \quad (f) \bigcap_{\alpha \in (0,\infty)} \left[1, 1 + \frac{1}{\alpha}\right].$$

1.2. Set theory and mathematical statements

As already mentioned, set theory provides a comfortable environment in which to do proofs and to learn the ins and outs of mathematical statements.[6] In this section we give a brief account of the various ways mathematical statements can be worded using the background of set theory.

1.2.1. More on "if ... then" statements. We begin by exploring different ways of saying "if ... then." Consider again the statement that $A \subseteq B$:
$$\text{If } x \in A, \text{ then } x \in B.$$
We can also write this as
$$x \in A \text{ implies } x \in B \qquad \text{or} \qquad x \in A \Longrightarrow x \in B;$$
that is, x belongs to A implies that x also belongs to B. Here, $\Longrightarrow$ is the common symbol for "implies". Another way to say this is
$$x \in A \text{ only if } x \in B;$$
that is, the object x belongs to A only if x also belongs to B. Here is yet one more way to write the statement:
$$x \in B \text{ if } x \in A;$$
that is, the object x belongs to B if, or given that, the object x belongs to A. Finally, we also know that the **contrapositive** statement says the same thing:
$$\text{If } x \notin B, \text{ then } x \notin A.$$
Thus, the following statements all mean the same thing:
$$\begin{aligned} &\text{If } x \in A, \text{ then } x \in B; \qquad x \in A \text{ implies } x \in B; \qquad \text{Given } x \in A,\ x \in B; \\ &x \in A \text{ only if } x \in B; \qquad x \in B \text{ if } x \in A; \qquad \text{If } x \notin B, \text{ then } x \notin A. \end{aligned} \tag{1.3}$$

We now consider each of these set statements in more generality. First of all, a "statement" in the mathematical sense is a statement that is either true or false, but never both; much in the same way that we work only with sets and objects such

[5] *The point of philosophy is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it. Bertrand Russell (1872–1970).*

[6] *Another advantage of a mathematical statement is that it is so definite that it might be definitely wrong; and if it is found to be wrong, there is a plenteous choice of amendments ready in the mathematicians' stock of formulae. Some verbal statements have not this merit; they are so vague that they could hardly be wrong, and are correspondingly useless. Lewis Fry Richardson (1881–1953). Mathematics of War and Foreign Politics.*

that any given object is either in or not in a given set, but never both. In a day when "there are no absolutes" is commonly taught in high school, it may take a while to fully grasp the language of mathematics. A mathematical statement always has **hypotheses** or **assumptions**, and a **conclusion**. Almost always there are **hidden assumptions**, that is, assumptions that are not stated, but taken for granted, because the context makes it clear what these assumptions are. Whenever you read a mathematical statement, make sure that you fully understand the hypotheses or assumptions (including hidden ones) and the conclusion. For the statement "If $x \in A$, then $x \in B$", the assumption is $x \in A$ and the conclusion is $x \in B$. The "if-then" wording means: If the assumptions ($x \in A$) are true, then the conclusion ($x \in B$) is also true, or stated another way, given that the assumptions are true, the conclusion follows. Let P denote the statement that $x \in A$ and Q the statement that $x \in B$. Then each of the following statements are equivalent, that is, the truth of any one statement implies the truth of any of the other statements:[7]

$$\boxed{\begin{array}{lll} \text{If } P\text{, then } Q; & P \text{ implies } Q; & \text{Given } P\text{, } Q \text{ holds;} \\ P \text{ only if } Q; & Q \text{ if } P; & \text{If not } Q\text{, then not } P. \end{array}} \tag{1.4}$$

Each of these statements are for P being $x \in A$ and Q being $x \in B$, but as you probably guess, they work for any mathematical statements P and Q. Let us consider statements concerning real numbers.

Example 1.14. Let P be the statement that $x > 5$. Let Q be the statement that $x^2 > 100$. Then each of the statements are equivalent:

$$\text{If } x > 5\text{, then } x^2 > 100; \qquad x > 5 \text{ implies } x^2 > 100; \qquad \text{Given } x > 5,\ x^2 > 100;$$

$$x > 5 \text{ only if } x^2 > 100; \qquad x^2 > 100 \text{ if } x > 5; \qquad \text{If } x^2 \le 100\text{, then } x \le 5.$$

The hidden assumptions are that x represents a real number and that the real numbers satisfy all the axioms you think they do. Of course, any one (and hence every one) of these six statements is false. For instance, $x = 6 > 5$ is true, but $x^2 = 36$, which is not greater than 100.

Example 1.15. Let P be the statement that $x^2 = 2$. Let Q be the statement that x is irrational. Then each of the statements are equivalent:

$$\text{If } x^2 = 2\text{, then } x \text{ is irrational;} \qquad x^2 = 2 \text{ implies } x \text{ is irrational;}$$

$$\text{Given } x^2 = 2,\ x \text{ is irrational;} \qquad x^2 = 2 \text{ only if } x \text{ is irrational;}$$

$$x \text{ is irrational if } x^2 = 2; \qquad \text{If } x \text{ is rational, then } x^2 \neq 2.$$

Again, the hidden assumptions are that x represents a real number and that the real numbers satisfy all their usual properties. Any one (and hence every one) of these six statements is of course true (since we are told since high school that $\pm\sqrt{2}$ are irrational; we shall prove this fact in Section 2.6).

As these two examples show, it is very important to remember that none of the statements in (1.4) assert that P or Q is true; they simply state *if* P is true, *then* Q is also true.

[7] P implies Q is sometimes translated as "P is sufficient for Q" in the sense that the truth of P is sufficient or enough or ample to imply that Q is also true. P implies Q is also translated "Q is necessary for P" because Q is necessarily true given that P is true. However, we shall not use this language in this book.

1.2.2. Converse statements and "if and only if" statements. Given a statement P implies Q, the reverse statement Q implies P is called the **converse** statement. For example, back to set theory, the converse of the statement

$$\text{If } x \in A, \text{ then } x \in B; \text{ that is, } A \subseteq B,$$

is just the statement that

$$\text{If } x \in B, \text{ then } x \in A; \text{ that is, } B \subseteq A.$$

These set theory statements make it clear that the converse of a true statement may not be true, for $\{e, \pi\} \subseteq \{e, \pi, i\}$, but $\{e, \pi, i\} \not\subseteq \{e, \pi\}$. Let us consider examples with real numbers.

Example 1.16. The statement "If $x^2 = 2$, then x is irrational" is true, but its converse statement, "If x is irrational, then $x^2 = 2$," is false.

Statements for which the converse is equivalent to the original statement are called "if and only if" statements.

Example 1.17. Consider the statement "If $x = -5$, then $2x + 10 = 0$." This statement is true. Its converse statement is "If $2x + 10 = 0$, then $x = -5$." By solving the equation $2x + 10 = 0$, we see that the converse statement is also true.

The implication $x = -5 \Longrightarrow 2x + 10 = 0$ can be written

$$2x + 10 = 0 \text{ if } x = -5, \tag{1.5}$$

while the implication $2x + 10 = 0 \Longrightarrow x = -5$ can be written

$$2x + 10 = 0 \text{ only if } x = -5. \tag{1.6}$$

Combining the two statements (1.5) and (1.6) into one statement, we get

$$2x + 10 = 0 \textit{ if } \text{and } \textit{only if } x = -5,$$

which is often denoted by a double arrow

$$2x + 10 = 0 \Longleftrightarrow x = -5,$$

or in more common terms, $2x + 10 = 0$ is equivalent to $x = -5$. We regard the statements $2x + 10 = 0$ and $x = -5$ as equivalent because if one statement is true, then so is the other one; hence the wording "is equivalent to". In summary, if both statements

$$Q \text{ if } P \text{ (that is, } P \Longrightarrow Q) \quad \text{and} \quad Q \text{ only if } P \text{ (that is, } Q \Longrightarrow P)$$

hold, then we write

$$Q \text{ if and only if } P \qquad \text{or} \qquad Q \Longleftrightarrow P.$$

Also, if you are asked to prove a statement "Q if and only if P", then you have to prove both the "if" statement "Q if P" (that is, $P \Longrightarrow Q$) and the "only if" statement "Q only if P" (that is, $Q \Longrightarrow P$).

The if and only if notation $\Longleftrightarrow$ comes in quite handy in proofs whenever we want to move from one statement to an equivalent one.

Example 1.18. Recall that in the proof of Theorem 1.3, we wanted to show that $A \cap \bigcup_\alpha A_\alpha = \bigcup_\alpha (A \cap A_\alpha)$, which means that $A \cap \bigcup_\alpha A_\alpha \subseteq \bigcup_\alpha (A \cap A_\alpha)$ and $\bigcup_\alpha (A \cap A_\alpha) \subseteq A \cap \bigcup_\alpha A_\alpha$; that is,

$$x \in A \cap \bigcup_\alpha A_\alpha \implies x \in \bigcup_\alpha (A \cap A_\alpha) \ \text{ and } \ x \in \bigcup_\alpha (A \cap A_\alpha) \implies x \in A \cap \bigcup_\alpha A_\alpha,$$

which is to say, we wanted to prove that

$$x \in A \cap \bigcup_\alpha A_\alpha \iff x \in \bigcup_\alpha (A \cap A_\alpha).$$

We can prove this quick and simple using $\iff$:

$$\begin{aligned} x \in A \cap \bigcup_\alpha A_\alpha \iff x \in A \text{ and } x \in \bigcup_\alpha A_\alpha &\iff x \in A \text{ and } x \in A_\alpha \text{ for some } \alpha \\ &\iff x \in A \cap A_\alpha \text{ for some } \alpha \\ &\iff x \in \bigcup_\alpha (A \cap A_\alpha). \end{aligned}$$

Just make sure that if you use $\iff$, the expression to the immediate left and right of $\iff$ are indeed equivalent.

1.2.3. Negations and logical quantifiers. We already know that a statement and its contrapositive are always equivalent: "if P, then Q" is equivalent to "if not Q, then not P". Therefore, it is important to know how to "not" something, that is, find the **negation**. Sometimes the negation is obvious.

Example 1.19. The negation of the statement that $x > 5$ is $x \leq 5$, and the negation of the statement that x is irrational is that x is rational. (In both cases, we are working under the unstated assumptions that x represents a real number.)

But some statements are not so easy especially when there are **logical quantifiers**: "for every" = "for all" (sometimes denoted by $\forall$ in class, but not in this book), and "for some" = "there exists" = "there is" = "for at least one" (sometimes denoted by $\exists$ in class, but not in this book). The equal signs represent the fact that we mathematicians consider "for every" as another way of saying "for all", "for some" as another way of saying "there exists", and so forth. Working under the assumptions that all numbers we are dealing with are real, consider the statement

$$\text{For every } x,\ x^2 \geq 0. \tag{1.7}$$

What is the negation of this statement? One way to find out is to think of this in terms of set theory. Let $A = \{x \in \mathbb{R}\,;\, x^2 \geq 0\}$. Then the statement (1.7) is just that $A = \mathbb{R}$. It is obvious that the negation of the statement $A = \mathbb{R}$ is just $A \neq \mathbb{R}$. Now this means that there must exist some real number x such that $x \notin A$. In order for x to not be in A, it must be that $x^2 < 0$. Therefore, $A \neq \mathbb{R}$ just means that there is a real number x such that $x^2 < 0$. Hence, the negation of (1.7) is just

$$\text{For at least one } x,\ x^2 < 0.$$

Thus, the "for every" statement (1.7) becomes a "there is" statement. In general, the negation of a statement of the form

"For every x, P" is the statement "For at least one x, not P."

Similarly, the negation of a "there is" statement becomes a "for every" statement. Explicitly, the negation of

"For at least one x, Q" is the statement "For every x, not Q."

For instance, with the understanding that x represents a real number, the negation of "There is an x such that $x^2 = 2$" is "For every x, $x^2 \neq 2$".

Exercises 1.2.

1. In this problem all numbers are understood to be real. Write down the contrapositive and converse of the following statement:
$$\text{If } x^2 - 2x + 10 = 25, \text{ then } x = 5,$$
and determine which (if any) of the three statements are true.
2. Write the negation of the following statements, where x represents an integer.
   (a) For every x, $2x + 1$ is odd.
   (b) There is an x such that $2^x + 1$ is prime.[8]
3. Here are some more set theory proofs to brush up on.
   (a) Prove that $(A^c)^c = A$.
   (b) Prove that $A = A \cup B$ if and only if $B \subseteq A$.
   (c) Prove that $A = A \cap B$ if and only if $A \subseteq B$.

1.3. What are functions?

In high school we learned that a function is a "rule that assigns to each input exactly one output". In practice, what usually comes to mind is a formula, such as
$$p(x) = x^2 - 3x + 10.$$
In fact, Leibniz who in 1692 (or as early as 1673) introduced the word "function" [**221**, p. 272] and to all mathematicians of the eighteenth century, a function was always associated to some type of analytic expression "a formula". However, because of necessity to problems in mathematical physics, the notion of function was generalized throughout the years and in this section we present the modern view of what a function is; see [**118**] or [**137, 138**] for some history.

1.3.1. (Cartesian) product. If A and B are sets, their **(Cartesian) product**, denoted by $A \times B$, is the set of all 2-tuples (or ordered pairs) where the first element is in A and the second element is in B. Explicitly,
$$\boxed{A \times B := \{(a, b)\,;\, a \in A,\ b \in B\}.}$$
We use the adjective "ordered" because we distinguish between ordered pairs, e.g. $(e, \pi) \neq (\pi, e)$, but as sets we regard then as equal, $\{e, \pi\} = \{\pi, e\}$. Of course, one can also define the product of any finite number of sets
$$A_1 \times A_2 \times \cdots \times A_m$$
as the set of all m-tuples $(a_1, \ldots, a_m)$ where $a_k \in A_k$ for each $k = 1, \ldots, m$.

Example 1.20. Of particular interest is m-dimensional Euclidean space
$$\mathbb{R}^m := \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{(m \text{ times})},$$
which is studied in Section 2.8.

[8] A number that is not prime is called composite.

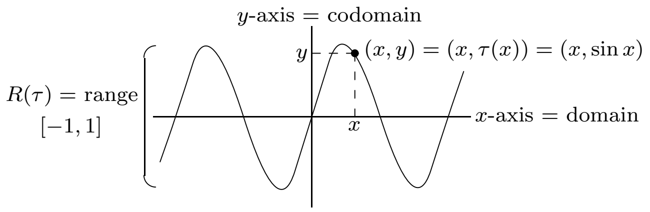


FIGURE 1.4. The function $\tau : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $\tau(x) = \sin x$.

1.3.2. Functions. Let X and Y be sets. Informally, we say that a function f from X into Y, denoted $f : X \longrightarrow Y$, is a rule that associates to each element $x \in X$, a single element $y \in Y$. Mathematically, a **function** from X into Y is a subset f of the product $X \times Y$ such that each element $x \in X$ appears exactly once as the first entry of an ordered pair in the subset f. Explicitly, for each $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$.

Example 1.21. For instance,

$$p = \{(x, y)\,;\, x \in [0, 1] \text{ and } y = x^2 - 3x + 10\} \subseteq [0, 1] \times \mathbb{R} \tag{1.8}$$

defines a function $p : [0, 1] \longrightarrow \mathbb{R}$, since p is an example of a subset of $[0, 1] \times \mathbb{R}$ such that each real number $x \in [0, 1]$ appears exactly once as the first entry of an ordered pair in p; e.g., the real number 1 appears as the first entry of $(1, 1^2 - 3\cdot 1 + 10) = (1, 8)$, and there is no other ordered pair in the set (1.8) with 1 as the first entry. Thus, p satisfies the mathematical definition of a function as you thought it should!

If $f : X \longrightarrow Y$ is a function, then we say that f **maps** X **into** Y. If $Y = X$ so that $f : X \longrightarrow X$, we say that f is a **function on** X. For a function $f : X \longrightarrow Y$, the **domain** of f is X, the **codomain** or **target** of f is Y, and the **range** of f, sometimes denoted $R(f)$, is the set of all elements in Y that occur as the second entry of an ordered pair in f. If $(x, y) \in f$ (recall that f is a set of ordered pairs), then we call the second entry y the **value** or **image** of the function at x and we write $y = f(x)$, and sometimes we write

$$x \mapsto y = f(x).$$

Using this $f(x)$ notation, which by the way was introduced in 1734 by Leonhard Euler (1707–1783) [**171**], [**36**, p. 443], we have

$$f = \{(x, y) \in X \times Y\,;\, y = f(x)\} \;=\; \{(x, f(x))\,;\, x \in X\} \;\subseteq\; X \times Y, \tag{1.9}$$

and

$$R(f) = \{y \in Y\,;\, y = f(x) \text{ for some } x \in X\} = \{f(x)\,;\, x \in X\}.$$

See Figure 1.4 for the familiar **graph** illustration of domain, codomain, and range for the trig function $\tau : \mathbb{R} \longrightarrow \mathbb{R}$ given by $\tau(x) = \sin x$. Also using this $f(x)$ notation, we can return to our previous ways of thinking of functions. For instance, we can say "let $p : [0, 1] \longrightarrow \mathbb{R}$ be the function $p(x) = x^2 - 3x + 10$" or "let $p : [0, 1] \longrightarrow \mathbb{R}$ be the function $x \mapsto x^2 - 3x + 10$", by which we mean of course the set (1.8). In many situations in this book, we are dealing with a fixed codomain; for example, with real-valued functions or stated another way, functions whose codomain is $\mathbb{R}$. Then we can omit the codomain and simply say, "let p be the function $x \mapsto x^2 - 3x + 10$ for $x \in [0, 1]$". In this case we again mean the set (1.8).

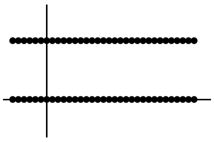


FIGURE 1.5. An attempted graph of Dirichlet's function.

We shall also deal quite a bit with complex-valued functions, that is, functions whose codomain is $\mathbb{C}$. Then if we say, "let f be a complex-valued function on $[0,1]$", we mean that $f : [0,1] \longrightarrow \mathbb{C}$ is a function. Here are some more examples.

Example 1.22. Consider the function $s : \mathbb{N} \longrightarrow \mathbb{R}$ defined by

$$s = \left\{ \left(n, \frac{(-1)^n}{n} \right) \; ; \, n \in \mathbb{N} \right\} \subseteq \mathbb{N} \times \mathbb{R}.$$

We usually denote $s(n) = \frac{(-1)^n}{n}$ by s_n and write $\{s_n\}$ for the function s, and we call $\{s_n\}$ a **sequence** of real numbers. We shall study sequences in great depth in Chapter 3.

Example 1.23. Here is a "piecewise" defined function: $a : \mathbb{R} \longrightarrow \mathbb{R}$,

$$a(x) = \begin{cases} x & \text{if } x \geq 0; \\ -x & \text{if } x < 0. \end{cases}$$

Of course, $a(x)$ is usually denoted by $|x|$ and is called the **absolute value function**.

Example 1.24. Here's an example of a "pathological function," the **Dirichlet function**, named after Johann Peter Gustav Lejeune Dirichlet (1805–1859), which is the function $D : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$D(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

This function was introduced in 1829 in Dirichlet's study of Fourier series and was the first function (1) not given by an analytic expression and (2) not continuous anywhere [**118**, p. 292]. See Section 4.3 for more on continuous functions. See Figure 1.5 for an attempted graph of Dirichlet's function.

In elementary calculus, you often encountered composition of functions when learning, for instance, the chain rule. Here is the precise definition of composition. If $f : X \longrightarrow Y$ and $g : Z \longrightarrow X$, then the **composition** $f \circ g$ is the function

$$f \circ g : Z \longrightarrow Y$$

defined by $(f \circ g)(z) := f(g(z))$ for all $z \in Z$. As a set of ordered pairs, $f \circ g$ is given by (do you see why?)

$$f \circ g = \{(z,y) \in Z \times Y \, ; \text{ for some } x \in X, \, (z,x) \in g \text{ and } (x,y) \in f\} \subseteq Z \times Y.$$

Also, when learning about the exponential or logarithmic functions, you probably encountered inverse functions. Here are some definitions related to this area. A function $f : X \longrightarrow Y$ is called **one-to-one** or **injective** if for each $y \in R(f)$, there is only one $x \in X$ with $y = f(x)$. Another way to state this is

$$\boxed{f \text{ is one-to-one means: If } f(x_1) = f(x_2), \text{ then } x_1 = x_2.} \tag{1.10}$$

In terms of the contrapositive, we have

$$\boxed{f \text{ is one-to-one means: If } x_1 \neq x_2, \text{ then } f(x_1) \neq f(x_2).} \tag{1.11}$$

In case $f : X \longrightarrow Y$ is injective, the **inverse** map f^{-1} is the map with domain $R(f)$ and codomain X:

$$f^{-1} : R(f) \longrightarrow X$$

defined by $f^{-1}(y) := x$ where $y = f(x)$. The function f is called **onto** or **surjective** if $R(f) = Y$; that is,

$$\boxed{f \text{ is onto means: For every } y \in Y \text{ there is an } x \in X \text{ such that } y = f(x).} \tag{1.12}$$

A one-to-one and onto map is called a **bijection**. Here are some examples.

Example 1.25. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by $f(x) = x^2$. Then f is not one-to-one because e.g. (see the condition (1.11)) $2 \neq -2$ yet $f(2) = f(-2)$. This function is also not onto because it fails (1.12): e.g. for $y = -1 \in \mathbb{R}$ there is no $x \in \mathbb{R}$ such that $-1 = f(x)$.

Example 1.26. In elementary calculus, we learn that the exponential function

$$\exp : \mathbb{R} \longrightarrow (0, \infty), \qquad f(x) = e^x,$$

is both one-to-one and onto, that is, a bijection, with inverse

$$\exp^{-1} : (0, \infty) \longrightarrow \mathbb{R}, \qquad f^{-1}(x) = \log x.$$

Here, $\log x$ denotes the "natural logarithm", which in many calculus courses is denoted by $\ln x$, with $\log x$ denoting the base 10 logarithm; however in this book and in most advanced math texts, $\log x$ denotes the natural logarithm. In Chapter 3 we shall define the exponential and logarithmic functions rigorously.

1.3.3. Images and inverse images. Functions act on sets as follows. Given a function $f : X \longrightarrow Y$ and a set $A \subseteq X$, we define

$$\boxed{f(A) := \{f(x)\,;\, x \in A\} = \{y \in Y\,;\, y = f(x) \text{ for some } x \in A\},}$$

and call this set the **image** of A under f. Thus,

$$y \in f(A) \quad \Longleftrightarrow \quad y = f(x) \text{ for some } x \in A.$$

Given a set $B \subseteq Y$, we define

$$\boxed{f^{-1}(B) := \{x \in X\,;\, f(x) \in B\},}$$

and call this set the **inverse image** or **preimage** of B under f. Thus,

$$x \in f^{-1}(B) \quad \Longleftrightarrow \quad f(x) \in B.$$

Warning: The notation f^{-1} in the preimage $f^{-1}(B)$ is only notation and *does not* represent the inverse function of f. (Indeed, the function may not have an inverse so the inverse function may not even be defined.)

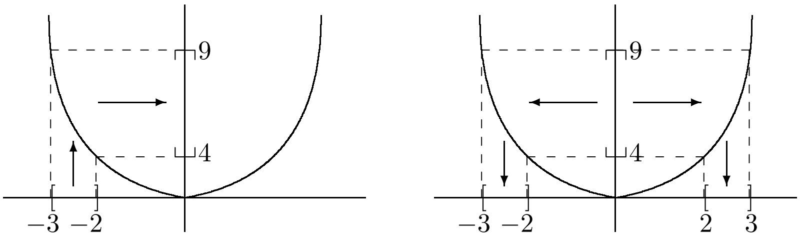


FIGURE 1.6. (Left-hand picture) The function $f(x) = x^2$ takes all the points in $[-3,-2]$ to the set $[4,9]$, so $f([-3,-2]) = [4,9]$. (Right-hand picture) $f^{-1}([4,9])$ consists of every point in $\mathbb{R}$ that f brings inside of $[4,9]$, so $f^{-1}([4,9]) = [-3,-2] \cup [2,3]$.

Example 1.27. Let $f(x) = x^2$ with domain and range in $\mathbb{R}$. Then as we can see in Figure 1.6,

$$f([-3,-2]) = [4,9] \quad \text{and} \quad f^{-1}([4,9]) = [-3,-2] \cup [2,3].$$

Here are more examples: You are invited to check that

$$f((1,2]) = (1,4], \quad f^{-1}([-4,-1)) = \varnothing, \quad f^{-1}((1,4]) = [-2,-1) \cup (1,2].$$

The following theorem gives the main properties of images and inverse images.

THEOREM 1.4. *Let $f : X \longrightarrow Y$, let $B, C \subseteq Y$, $\{B_\alpha\}$ be a family of subsets of Y, and let $\{A_\alpha\}$ a family of subsets of X. Then*

$$f^{-1}(C \setminus B) = f^{-1}(C) \setminus f^{-1}(B), \qquad f^{-1}\Big(\bigcup_\alpha B_\alpha\Big) = \bigcup_\alpha f^{-1}(B_\alpha),$$
$$f^{-1}\Big(\bigcap_\alpha B_\alpha\Big) = \bigcap_\alpha f^{-1}(B_\alpha), \qquad f\Big(\bigcup_\alpha A_\alpha\Big) = \bigcup_\alpha f(A_\alpha).$$

PROOF. Using the definition of inverse image and set difference, we have

$$\begin{aligned} x \in f^{-1}(C \setminus B) \Longleftrightarrow f(x) \in C \setminus B \Longleftrightarrow & f(x) \in C \text{ and } f(x) \notin B \\ \Longleftrightarrow & x \in f^{-1}(C) \text{ and } x \notin f^{-1}(B) \\ \Longleftrightarrow & x \in f^{-1}(C) \setminus f^{-1}(B). \end{aligned}$$

Thus, $f^{-1}(C \setminus B) = f^{-1}(C) \setminus f^{-1}(B)$.

Using the definition of inverse image and union, we have

$$\begin{aligned} x \in f^{-1}\Big(\bigcup_\alpha B_\alpha\Big) \Longleftrightarrow f(x) \in \bigcup_\alpha B_\alpha \Longleftrightarrow & f(x) \in B_\alpha \text{ for some } \alpha \\ \Longleftrightarrow & x \in f^{-1}(B_\alpha) \text{ for some } \alpha \\ \Longleftrightarrow & x \in \bigcup_\alpha f^{-1}(B_\alpha). \end{aligned}$$

Thus, $f^{-1}\Big(\bigcup_\alpha B_\alpha\Big) = \bigcup_\alpha f^{-1}(B_\alpha)$. The proof of the last two properties in this theorem are similar enough to the proof just presented that we leave their verification to the reader. □

We end this section with some definitions needed for the exercises. Let X be a set and let A be any subset of X. The **characteristic function** of A is the function $\chi_A : X \longrightarrow \mathbb{R}$ defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{if } x \notin A. \end{cases}$$

The sum and product of two characteristic function χ_A and χ_B are the functions $\chi_A + \chi_B : X \longrightarrow \mathbb{R}$ and $\chi_A \cdot \chi_B : X \longrightarrow \mathbb{R}$ defined by

$$(\chi_A+\chi_B)(x) = \chi_A(x)+\chi_B(x) \quad \text{and} \quad (\chi_A\cdot\chi_B)(x) = \chi_A(x)\cdot\chi_B(x), \quad \text{for all } x \in X.$$

Of course, the sum and product of *any* functions $f : X \longrightarrow \mathbb{R}$ and $g : X \longrightarrow \mathbb{R}$ are defined in the same way. We can also replace $\mathbb{R}$ by, say $\mathbb{C}$, or by any set Y as long as "+" and "·" are defined on Y. Given any constant $c \in \mathbb{R}$, we denote by the same letter the function $c : X \longrightarrow \mathbb{R}$ defined by $c(x) = c$ for all $x \in X$. This is the **constant function** c. For instance, 0 is the function defined by $0(x) = 0$ for all $x \in X$. The **identity map** on X is the map defined by $I(x) = x$ for all $x \in X$. Finally, we say that two functions $f : X \longrightarrow Y$ and $g : X \longrightarrow Y$ are **equal** if $f = g$ as subsets of $X \times Y$, which holds if and only if $f(x) = g(x)$ for all $x \in X$.

Exercises 1.3.

1. Which of the following subsets of $\mathbb{R} \times \mathbb{R}$ define functions from $\mathbb{R}$ to $\mathbb{R}$?

$$(a)\ A_1 = \{(x,y) \in \mathbb{R}\times\mathbb{R}\,;\, x^2 = y\}, \quad (b)\ A_2 = \{(x,y) \in \mathbb{R}\times\mathbb{R}\,;\, x = \sin y\},$$
$$(c)\ A_3 = \{(x,y) \in \mathbb{R}\times\mathbb{R}\,;\, y = \sin x\}, \quad (d)\ A_4 = \{(x,y) \in \mathbb{R}\times\mathbb{R}\,;\, x = 4y-1\}.$$

   (Assume well-known properties of trig functions.) Of those sets which do define functions, find the range of the function. Is the function is one-to-one; is it onto?

2. Let $f(x) = 1 - x^2$. Find

$$f([1,4]), \quad f([-1,0] \cup (2,10)), \quad f^{-1}([-1,1]), \quad f^{-1}([5,10]), \quad f(\mathbb{R}), \quad f^{-1}(\mathbb{R}).$$

3. If $f : X \longrightarrow Y$ and $g : Z \longrightarrow X$ are bijective, prove that $f \circ g$ is a bijection and $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.
4. Let $f : X \longrightarrow Y$ be a function.
   (a) Given any subset $B \subseteq Y$, prove that $f(f^{-1}(B)) \subseteq B$.
   (b) Prove that $f(f^{-1}(B)) = B$ for all subsets B of Y if and only if f is surjective.
   (c) Given any subset $A \subseteq X$, prove that $A \subseteq f^{-1}(f(A))$.
   (d) Prove that $A = f^{-1}(f(A))$ for all subsets A of X if and only if f is injective.
5. Let $f : X \longrightarrow Y$ be a function. Show that f is one-to-one if and only if there is a function $g : Y \longrightarrow X$ such that $g \circ f$ is the identity map on X. Show that f is onto if and only if there is a function $h : Y \longrightarrow X$ such that $f \circ h$ is the identity map on Y.
6. (Cf. [**152**]) In this problem we give various applications of characteristic functions to prove statements about sets. First, prove at least two of (a) – (e) of the following. (a) $\chi_X = 1$, $\chi_\varnothing = 0$; (b) $\chi_A \cdot \chi_B = \chi_B \cdot \chi_A = \chi_{A\cap B}$ and $\chi_A \cdot \chi_A = \chi_A$; (c) $\chi_{A\cup B} = \chi_A + \chi_B - \chi_A \cdot \chi_B$; (d) $\chi_{A^c} = 1 - \chi_A$; (e) $\chi_A = \chi_B$ if and only if $A = B$. Here are some applications of these properties Prove the distributive law:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

   by showing that the characteristic functions of each side are equal as functions. Then invoke (e) to demonstrate equality of sets. Prove the nonobvious equality

$$(A \cap B^c) \cap (C^c \cap A) = A \cap (B \cup C)^c.$$

   Here's a harder question: Consider the sets $(A \cup B) \cap C$ and $A \cup (B \cap C)$. When, if ever, are they equal? When is one set a subset of the other?

CHAPTER 2

Numbers, numbers, and more numbers

I believe there are 15,747,724,136,275,002,577,605,653,961,181,555,468, 044,717,914,527,116,709,366,231,425,076,185,631,031,296 protons in the universe and the same number of electrons.
Sir Arthur Eddington (1882–1944), "The Philosophy of Physical Science". Cambridge, 1939.

This chapter is on the study of numbers. Of course, we all have a working understanding of the real numbers and we use many aspects of these numbers in everyday life: tallying up tuition and fees, figuring out how much we have left on our food cards, etc. We have accepted from our childhood all the properties of numbers that we use everyday. In this chapter we shall actually *prove* these properties.

In everyday life, what usually comes to mind when we think of "numbers" are the counting, or natural, numbers $1, 2, 3, 4, \ldots$. We shall study the natural numbers and their properties in Sections 2.1 and 2.2. These numbers have been used from the beginning. Later, the Hindus became the first to systematically use "zero" and "negative" integers [**35**], [**36**, p. 220]; for example, Brahmagupta (598–670) gave arithmetic rules for multiplying and dividing with such numbers (although he mistakenly believed that $0/0 = 0$). We study the integers in Sections 2.3, 2.4, and 2.5. Everyday life forces us to talk about fractions, for example, $2/3$ of a pizza "two pieces of a pizza divided into three equal parts". Such fractions (and their negatives and zero) make up the so-called rational numbers, which are called *rational* not because they are "sane" or "comprehensible", but simply because they are *ratios* of integers. It was a shock to the Greeks who discovered that the rational numbers are not enough to describe nature. They noticed that according to the Pythagorean theorem, the length of the hypotenuse of a triangle with sides of length 1 is $\sqrt{1^2 + 1^2} = \sqrt{1 + 1} = \sqrt{2}$. We shall prove that $\sqrt{2}$ is "irrational", which simply means "not rational," that is, not a ratio of integers. In fact, we'll see that "most" numbers that you encountered in high school:

> *Square roots and more generally n-th roots, roots of polynomials, and values of trigonometric and logarithmic functions, are mostly irrational!*

You'll have to wait for this mouth-watering subject until Section 2.6! In Section 2.7 we study the all-important property of the real numbers called the completeness property, which in some sense says that real numbers can describe any length whatsoever. In Sections 2.8 and 2.9, we leave the one-dimensional real line and discuss m-dimensional space and the complex numbers (which is really just two-dimensional space). Finally, in Section 2.10 we define "most" using cardinality and show that "most" real numbers are not only irrational, they are transcendental.

Chapter 2 objectives: The student will be able to ...

- state the fundamental axioms of the natural, integer, and real number systems.

- Explain how the completeness axiom of the real number system distinguishes this system from the rational number system in a powerful way.
- prove statements about numbers from basic axioms including induction.
- Define $\mathbb{R}^m$ and $\mathbb{C}$ and the norms on these spaces.
- Explain cardinality and how "most" real numbers are irrational or even transcendental.

2.1. The natural numbers

The numbers we encounter most often in "everyday life" are the counting numbers, or the positive whole numbers

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \ldots$$

These are called the **natural numbers** and have been used since the beginning. In this section we study the essential properties of these fundamental numbers.

2.1.1. Axioms for natural numbers.

The *set*, or collection, of natural numbers is denoted by $\mathbb{N}$. We all know that if two natural numbers are added, we obtain a natural number; for example, $3 + 4 = 7$. Similarly, if two natural numbers are multiplied we get a natural number. We say that the natural numbers are **closed** under addition and multiplication. Thus, if a, b are in $\mathbb{N}$, then using the familiar notation for addition and multiplication, $a+b$ and $a \cdot b$ are also in $\mathbb{N}$.[1] The following properties of $+$ and $\cdot$ are also familiar.

Addition satisfies

(A1) $a + b = b + a$; (commutative law)
(A2) $(a + b) + c = a + (b + c)$. (associative law)

By the associative law, we may "drop" parentheses in sums of more than two numbers:

$$a + b + c \text{ is unambiguously defined as } (a + b) + c = a + (b + c).$$

Multiplication satisfies

(M1) $a \cdot b = b \cdot a$; (commutative law)
(M2) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$; (associative law)
(M3) there is a natural number denoted by 1 "one" such that

$$1 \cdot a = a = a \cdot 1. \quad \text{(existence of multiplicative identity)}$$

By the associative law for multiplication, we may "drop" parentheses:

$$a \cdot b \cdot c \text{ is unambiguously defined as } (a \cdot b) \cdot c = a \cdot (b \cdot c).$$

Addition and multiplication are related by

(D) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$. (distributive law)

As usual, we sometimes drop the dot $\cdot$ and just write ab for $a \cdot b$. The natural numbers are also ordered in the sense that you can compare the magnitude of any two of them; for example $2 < 5$ because five is three greater than two or two is three less than five. This inequality relationship satisfies the following familiar properties. Given any natural numbers a and b *exactly one* of the (in)equalities hold:

(O1) $a = b$;
(O2) $a < b$, which by definition means that $b = a + c$ for some natural number c;

[1] By the way, $+$ and $\cdot$ are functions, as we studied in Section 1.3, from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{N}$. These are not arbitrary functions but must satisfy the properties **(A)**, **(M)**, and **(D)** listed.

(O3) $b < a$, which by definition means that $a = b + c$ for some natural number c.

Thus, $2 < 5$ because $5 = 2 + c$ where $c = 3$. Of course, we write $a \leq b$ if $a < b$ or $a = b$. There are similar meanings for the opposite inequalities "$>$" and "$\geq$". The inequality signs $<$ and $>$ are called **strict**. There is one more property of the natural numbers called **induction**. Let M be a subset of $\mathbb{N}$.

(I) Suppose that M contains 1 and that M has the following property: If n belongs to M, then $n + 1$ also belongs to M. Then M contains all natural numbers.

The statement that $M = \mathbb{N}$ is "obvious" with a little thought. M contains 1. Because 1 belongs to M, by (**I**), we know that $1+1 = 2$ also belongs to M. Because 2 belongs to M, by (**I**) we know that $2 + 1 = 3$ also belongs to M. Assuming we can continue this process indefinitely makes it clear that $M = \mathbb{N}$.

Everyday experience convinces us that the counting numbers satisfy properties (**A**), (**M**), (**D**), (**O**), and (**I**). However, mathematically we will assume, or take by faith, the existence of a set $\mathbb{N}$ with operations $+$ and $\cdot$ that satisfy properties (**A**), (**M**), (**D**), (**O**), and (**I**).[2] From these properties alone we shall *prove* many well-known properties of these numbers that we have accepted since grade school. It is quite satisfying to see that many of the well-known properties about numbers that are memorized (or even those that are not so well-known) can in fact be proven from a basic set of axioms! The "rules of the game" to prove such properties is that we are allowed to prove statements only using facts that we already know are true either because these facts were given to us in a set of axioms, or because these facts have already been proven by us in this book, by your teacher in class, or by you in an exercise.

2.1.2. Proofs of well-known high school rules. Again, you are going to learn the language of proofs in the same way that a child learns to talk; by observing others prove things and imitating them, and eventually you will get the hang of it.

We begin by proving the familiar transitive law.

Theorem 2.1 (**Transitive law**). *If $a < b$ and $b < c$, then $a < c$.*

Proof. Suppose that $a < b$ and $b < c$. Then by definition of less than (recall the inequality law (**O2**) in Section 2.1.1), there are natural numbers d and e such that $b = a + d$ and $c = b + e$. Hence, by the associative law,

$$c = b + e = (a + d) + e = a + (d + e).$$

Thus, $c = a + f$ where $f = d + e \in \mathbb{N}$, so $a < c$ by definition of less than. □

Before moving on, we briefly analyze this theorem in view of what we learned in Section 1.2. The **hypotheses** or **assumptions** of this theorem are that a, b, and c are natural numbers with $a < b$ and $b < c$ and the **conclusion** is that $a < c$. Note that the fact that a, b, and c are natural numbers and that natural numbers are assumed to satisfy all their arithmetic and order properties were left unwritten in the statement of the proposition since these assumptions were understood within the context of this section. The "if-then" wording means: If the assumptions are true, then the conclusion is also true or given that the assumptions are true, the conclusion follows. We can also reword Theorem 2.1 as follows:

$$a < b \text{ and } b < c \text{ implies (also written } \Longrightarrow) \ a < c;$$

[2] Taking the axioms of set theory by faith, which we are doing in this book even though we haven't listed many of them(!), we can define the natural numbers as sets, see [**91**, Sec. 11].

that is, the truth of the assumptions implies the truth of the conclusion. We can also state this theorem as follows:

$$a < b \text{ and } b < c \text{ only if } a < c;$$

that is, the hypotheses $a < b$ and $b < c$ hold only if the conclusion $a < c$ also holds, or

$$a < c \text{ if } a < b \text{ and } b < c;$$

that is, the conclusion $a < c$ is true if, or given that, the hypotheses $a < b$ and $b < c$ are true. The kind of proof used in Theorem 2.1 is called a **direct proof**, where we take the hypotheses $a < b$ and $b < c$ as true and prove that the conclusion $a < c$ is true. We shall see other types of proofs later. We next give another easy and direct proof of the so-called "FOIL law" of multiplication. However, before proving this result, we note that the distributive law (**D**) holds from the right:

$$(a + b) \cdot c = ac + bc.$$

Indeed,

$$\begin{aligned} (a + b) \cdot c &= c \cdot (a + b) && \text{commutative law} \\ &= (c \cdot a) + (c \cdot b) && \text{distributive law} \\ &= (a \cdot c) + (b \cdot c) && \text{commutative law.} \end{aligned}$$

Theorem 2.2 (**FOIL law**). *For any natural numbers a, b, c, d, we have*

$$(a + b) \cdot (c + d) = ac + ad + bc + bd, \quad \textit{(first + outside + inside + last).}$$

Proof. We simply compute:

$$\begin{aligned} (a + b) \cdot (c + d) &= (a + b) \cdot c + (a + b) \cdot d && \text{distributive law} \\ &= (ac + bc) + (ad + bd) && \text{distributive law (from right)} \\ &= ac + (bc + (ad + bd)) && \text{associative law} \\ &= ac + ((bc + ad) + bd) && \text{associative law} \\ &= ac + ((ad + bc) + bd) && \text{commutative law} \\ &= ac + ad + bc + bd, \end{aligned}$$

where at the last step we dropped parentheses as we know we can in sums of more than two numbers (consequence of the associative law). □

We now prove the familiar cancellation properties of high school algebra.

Theorem 2.3. *Given any natural numbers a, b, c, we have*

$$a + c = b + c \quad \textit{if and only if} \quad a = b.$$

In particular, given $a + c = b + c$, we can "cancel" c, obtaining $a = b$.

Proof. Suppose that $a = b$, then because a and b are just different letters for the same natural number, we have $a + c = b + c$.

We now have to prove that if $a + c = b + c$, then $a = b$. To prove this, we use a **proof by contraposition**. This is how it works. We need to prove that if the assumption "$P : a + c = b + c$" is true, then the conclusion "$Q : a = b$" is also true. Instead, we shall prove the logically equivalent **contrapositive** statement: If the conclusion Q is false, then the assumption P must also false. The statement that Q

is false is just that $a \neq b$ and the statement that P is false is just that $a+c \neq b+c$. Thus, we must prove

$$\text{if } a \neq b, \text{ then } a+c \neq b+c.$$

To this end, assume that $a \neq b$; then either $a < b$ or $b < a$. Because the notation is entirely symmetric between a and b, we may presume that $a < b$. Then by definition of less than, we have $b = a + d$ for some natural number d. Hence, by the associative and commutative laws,

$$b + c = (a + d) + c = a + (d + c) = a + (c + d) = (a + c) + d.$$

Thus, by definition of less than, $a + c < b + c$, so $a + c \neq b + c$.

□

There is a multiplicative cancellation as well; see Problem 5b. Other examples of using the fundamental properties (**A**), (**M**), (**D**), and (**O**) of the natural numbers are found in the exercises. We now concentrate on the induction property (**I**).

2.1.3. Induction. We all know that every natural number is greater than or equal to one. Here is a proof!

THEOREM 2.4. *Every natural number is greater than or equal to one.*

PROOF. Rewording this as an "if-then" statement, we need to prove that if n is a natural number, then $n \geq 1$. To prove this, let $M = \{n \in \mathbb{N}\,;\, n \geq 1\}$, the collection all natural numbers greater than or equal to one. Then M contains 1. If a natural number n belongs to M, then by definition of M, $n \geq 1$. This means that $n = 1$ or $n > 1$. In the first case, $n + 1 = 1 + 1$, so by definition of less than, $1 < n + 1$. In the second case, $n > 1$ means that $n = 1 + m$ for some $m \in \mathbb{N}$, so $n + 1 = (1 + m) + 1 = 1 + (m + 1)$. Again by definition of less than, $1 < n + 1$. In either case, $n + 1$ also belongs to M. Thus by induction, $M = \mathbb{N}$. □

Now we prove the Archimedean ordering property of the natural numbers.

THEOREM 2.5 (**Archimedean ordering of** $\mathbb{N}$). *Given any natural numbers a and b there is a natural number n so that $b < a \cdot n$.*

PROOF. Let $a, b \in \mathbb{N}$; we need to *produce* an $n \in \mathbb{N}$ such that $b < a \cdot n$. By the previous theorem, either $a = 1$ or $a > 1$. If $a = 1$, then we set $n = b + 1$, in which case $b < b + 1 = 1 \cdot n$. If $1 < a$, then we can write $a = 1 + c$ for some natural number c. In this case, let $n = b$. Then,

$$a \cdot n = (1 + c) \cdot b = b + c \cdot b > b.$$

□

The following theorem contains an important property of the natural numbers. Its proof is an example of a **proof by contradiction** or **reductio ad absurdum**, whereby we start with a tentative assumption that the conclusion is false and then proceed with our argument until we eventually get a logical absurdity.

THEOREM 2.6 (**Well-ordering (principle) of** $\mathbb{N}$). *Any nonempty set of natural numbers has a smallest element; that is, an element less than or equal to any other member of the set.*

Proof. We need to prove that if A is a nonempty set of natural numbers, then A contains a natural number a so that $a < a'$ for any other element a' of A. Well, A either has this property or not; suppose, for the sake of contradiction, that A does not have a smallest element. From this assumption we shall derive a nonsense statement. Let $M = \{n \in \mathbb{N}\,;\, n < a \text{ for all } a \in A\}$. Note that since a natural number is never less than itself, M does not contain any element of A. In particular, since A is nonempty, M does not consist of all natural numbers. However, we shall prove by induction that M is all of $\mathbb{N}$. This of course would be a contradiction, for we already know that M is not all of $\mathbb{N}$.

To arrive at our contradiction, we first show that M contains 1. By Theorem 2.4, we know that 1 is less than or equal to every natural number; in particular, 1 is less than or equal to every element of A. Hence, if 1 is in A, then 1 would be the smallest element of A. However, we are assuming that A does not have a smallest element, so 1 cannot be in A. Hence, 1 is less than every element of A, so M contains 1.

Suppose that M contains n; we shall prove that M contains $n + 1$, that is, $n + 1$ is less than every element of A. Now either $n + 1 \in A$ or $n + 1 \notin A$. Suppose that $n + 1 \in A$ and let a be any element of A not equal to $n + 1$. Since $n < a$ (as $n \in M$), we can write $a = n + c$ for some natural number c. Note that $c \neq 1$ since by assumption $a \neq n + 1$. Thus (by Theorem 2.4) $c > 1$ and so we can write $c = 1 + d$ for some natural number d. Hence,

$$a = n + c = n + 1 + d,$$

which shows that $n + 1 < a$. This implies that $n + 1$ is the smallest element of A, which we know cannot exist. Hence, our supposition that $n + 1 \in A$ must have been incorrect, so $n + 1 \notin A$. In this case, the exact same argument just explained shows that $n + 1 < a$ for every element $a \in A$. Thus, $n + 1 \in M$, so by induction $M = \mathbb{N}$, and we arrive at our desired contradiction.

□

Finally, we remark that the letter 2 denotes, by definition, the natural number $1 + 1$. Since $2 = 1 + 1$, $1 < 2$ by definition of less than. The natural number 3 denotes the number $2 + 1 = 1 + 1 + 1$. By definition of less than, $2 < 3$. Similarly 4 is the number $3 + 1$, and so forth. Continuing in this manner we can assign the usual symbols to the natural numbers that we are accustomed to in "everyday life". In Problem 4 we see that there is no natural number between n and $n + 1$, so the sequence of symbols defined will cover all possible natural numbers.

All the letters in the following exercises represent natural numbers. In the following exercises, you are only allowed to use the axioms and properties of the natural numbers established in this section. Remember that if you can't see how to prove something after some effort, take a break (e.g. take a bus ride somewhere) and come back to the problem later.[3]

Exercises 2.1.

1. Prove that any natural number greater than 1 can be written in the form $m + 1$ where m is a natural number.

[3] *I entered an omnibus to go to some place or other. At that moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with non-Euclidean geometry. Henri Poincaré (1854–1912).*

2. Are there natural numbers a and b such that $a = a + b$? What logical inconsistency happens if such an equation holds?
3. Prove the following statements.
   (a) If $n^2 = 1$ (that is, $n \cdot n = 1$), then $n = 1$.
   (b) There does not exist a natural number n such that $2n = 1$.
   (c) There does not exist a natural number n such that $2n = 3$.
4. Prove the following statements.
   (a) If $n \in \mathbb{N}$, then there is no $m \in \mathbb{N}$ such that $n < m < n + 1$.
   (b) If $n \in \mathbb{N}$, then there is a unique $m \in \mathbb{N}$ satisfying $n < m < n + 2$; in fact, prove that the only such natural number is $m = n+1$. (That is, prove that $n+1$ satisfies the inequality and if m also satisfies the inequality, then $m = n + 1$.)
5. Prove the following statements.
   (a) $(a + b)^2 = a^2 + 2ab + b^2$, where a^2 means $a \cdot a$ and b^2 means $b \cdot b$.
   (b) For any fixed natural number c,
   $$a = b \text{ if and only if } a \cdot c = b \cdot c\ .$$
   Conclude that 1 is the only multiplicative identity (that is, if $a \cdot c = c$ for some $a, c \in \mathbb{N}$, then $a = 1$).
   (c) For any fixed natural number c,
   $$a < b \text{ if and only if } a + c < b + c\ .$$
   Also prove that
   $$a < b \text{ if and only if } a \cdot c < b \cdot c\ .$$
   (d) If $a < b$ and $c < d$, then $a \cdot c < b \cdot d$.
6. Let A be a finite collection of natural numbers. Prove that A has a largest element, that is, A contains a number n such that $n \geq m$ for every element m in A.
7. Many books take the well-ordering property as a fundamental axiom instead of the induction axiom. Replace the induction axiom by the well-ordering property.
   (a) Prove Theorem 2.4 using well-ordering. Suggestion: By well-ordering, $\mathbb{N}$ has a least element, call it n. We need to prove that $n \geq 1$. Assume that $n < 1$ and find another natural number less than n to derive a contradiction.
   (b) Prove the induction property.

2.2. The principle of mathematical induction

We now turn to the principle of mathematical induction and we give many applications of its use.

2.2.1. Principle of mathematical induction. The induction axiom of the natural numbers is the basis for the **principle of mathematical induction**, which goes as follows. Suppose that we are given a list of statements:

$$P_1, P_2, P_3, P_4, P_5, \ldots,$$

and suppose that (1) P_1 is true and (2) if n is a natural number and the statement P_n happens to be valid, then the statement P_{n+1} is also valid. Then it must be that every statement $P_1, P_2, P_3, \ldots$ is true. To see why every statement P_n is true, let M be the collection of all natural numbers n such that P_n is true. Then by (1), M contains 1 and by (2), if M contains a natural number n, then it contains $n + 1$. By the induction axiom, M must be all of $\mathbb{N}$; that is, P_n is true for every n. Induction is like dominoes: Line up infinitely many dominos in a row and knock down the first domino (that is, P_1 is true) and if the n-th domino knocks down the $(n + 1)$-st domino (that is, $P_n \Longrightarrow P_{n+1}$), then *every* domino gets knocked down

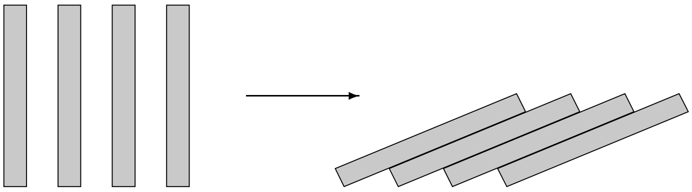


FIGURE 2.1. Induction is like dominoes.

(that is, all the statements $P_1, P_2, P_3, \ldots$ are true). See Figure 2.1 for a visual of this concept.

We now illustrate this principle through some famous examples. In order to present examples that have applicability in the sequel, we have to go outside the realm of natural numbers and assume basic familiarity with integers, real, and complex numbers. Integers will be discussed in the next section, real numbers in Sections 2.6 and 2.7, and complex numbers in Section 2.9.

2.2.2. Inductive definitions: Powers and sums. We of course know what 7^3 is, namely $7 \cdot 7 \cdot 7$. In general, we define a^n where a is any complex number called the **base** and n is a positive integer called the **exponent** as follows:

$$a^n := \underbrace{a \cdot a \cdots a}_{n \text{ times}}.$$

(Recall that ":=" means "equals by definition".)

Example 2.1. We can also define a^n using induction. Let P_n denote the statement "the power a^n is defined". We define $a^1 := a$. Assume that a^n has been defined. Then we define $a^{n+1} := a^n \cdot a$. Thus, the statement P_{n+1} is defined, so by induction a^n is defined for any natural number n.

Example 2.2. Using induction, we prove that for any natural numbers m and n, we have

$$a^{m+n} = a^m \cdot a^n. \tag{2.1}$$

Indeed, let us fix the natural number m and let P_n be the statement "Equation (2.1) holds for the natural number n". Certainly

$$a^{m+1} = a^m \cdot a = a^m \cdot a^1$$

holds by definition of a^{m+1}. Assume that (2.1) holds for a natural number n. Then by definition of the power and our induction hypothesis,

$$a^{m+(n+1)} = a^{(m+n)+1} = a^{m+n} \cdot a = a^m \cdot a^n \cdot a = a^m \cdot a^{n+1},$$

which is exactly the statement P_{n+1}. If $a \neq 0$ and we also define $a^0 := 1$, then as the reader can readily check, (2.1) continues to hold even if m or n is zero.

In elementary calculus, we were introduced to the summation notation. Let $a_0, a_1, a_2, a_3, \ldots$ be any list of complex numbers. For any natural number n, we define $\sum_{k=0}^n a_k$ as the sum of the numbers $a_0, \ldots, a_n$:

$$\sum_{k=0}^n a_k := a_0 + a_1 + \cdots + a_n.$$

By the way, in 1755 Euler introduced the sigma notation $\sum$ for summation [**171**].

Example 2.3. We also can define summation using induction. We define $\sum_{k=0}^{0} a_k := a_0$. For a natural number n, let P_n represent the statement "$\sum_{k=0}^{n} a_k$ is defined". We define

$$\sum_{k=0}^{1} a_k := a_0 + a_1.$$

Suppose that P_n holds for $n \in \mathbb{N}$; that is, $\sum_{k=0}^{n} a_k$ is defined. Then we define

$$\sum_{k=0}^{n+1} a_k := \left(\sum_{k=0}^{n} a_k \right) + a_{n+1}.$$

Thus, P_{n+1} holds. We conclude that the sum $\sum_{k=0}^{n} a_k$ is defined for $n = 0$ and for any natural number n.

2.2.3. Classic examples: The arithmetic and geometric progressions.

Example 2.4. First, we shall prove that for every natural number n, the sum of the first n integers is $n(n+1)/2$; that is,

$$\boxed{1 + 2 + \cdots + n = \frac{n(n+1)}{2}.} \tag{2.2}$$

Here, P_n represents the statement "Equation (2.2) holds". Certainly,

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

Thus, our statement is true for $n = 1$. Suppose our statement holds for a number n. Then adding $n + 1$ to both sides of (2.2), we obtain

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \frac{n(n+1)}{2} + (n+1) \\ &= \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+1+1)}{2}, \end{aligned}$$

which is exactly the statement P_{n+1}. Hence, by the principle of mathematical induction, every single statement P_n is true.

We remark that the high school way to prove P_n is to write the sum of the first n integers forward and backwards:

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

and

$$S_n = n + (n-1) + \cdots + 2 + 1.$$

Notice that the sum of each column is just $n+1$. Since there are n columns, adding these two expressions, we obtain $2S_n = n(n+1)$, which implies our result.

What if we only sum the odd integers? We get (proof left to you!)

$$1 + 3 + 5 + \cdots + (2n-1) = n^2.$$

Do you see why Figure 2.2 makes this formula "obvious"?

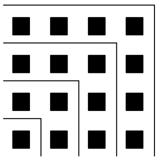


Figure 2.2. Sum of the first n odd numbers.

Example 2.5. We now consider the sum of a geometric progression. Let $a \neq 1$ be any complex number. We prove that for every natural number n,

$$\boxed{1 + a + a^2 + \cdots + a^n = \frac{1 - a^{n+1}}{1 - a}.} \tag{2.3}$$

Observe that

$$\frac{1 - a^2}{1 - a} = \frac{(1 + a)(1 - a)}{1 - a} = 1 + a,$$

so our assertion holds for $n = 1$. Suppose that the sum (2.3) holds for the number n. Then adding a^{n+1} to both sides of (2.3), we obtain

$$\begin{aligned} 1 + a + a^2 + \cdots + a^n + a^{n+1} &= \frac{1 - a^{n+1}}{1 - a} + a^{n+1} \\ &= \frac{1 - a^{n+1} + a^{n+1} - a^{n+2}}{1 - a} = \frac{1 - a^{n+2}}{1 - a}, \end{aligned}$$

which is exactly the equation (2.3) for $n + 1$. The completes the proof for the sum of a geometric progression.

The high school way to establish the sum of a geometric progression is to multiply

$$G_n = 1 + a + a^2 + \cdots + a^n$$

by a:

$$a\,G_n = a + a^2 + a^3 + \cdots + a^{n+1},$$

and then to subtract this equation from the preceding one and cancelling like terms:

$$(1 - a)G_n = G_n - a\,G_n = (1 + a + \cdots + a^n) - (a + \cdots + a^{n+1}) = 1 - a^{n+1}.$$

Dividing by $1 - a$ proves (2.3). Splitting the fraction at the end of (2.3) and solving for $1/(1 - a)$, we obtain the following version of the geometric progression

$$\boxed{\frac{1}{1 - a} = 1 + a + a^2 + \cdots + a^n + \frac{a^{n+1}}{1 - a}.} \tag{2.4}$$

2.2.4. More sophisticated examples. Here's a famous inequality due to Jacob (Jacques) Bernoulli (1654–1705) that we'll have to use on many occasions.

Theorem 2.7 (**Bernoulli's inequality**). *For any real number $a > -1$ and any natural number n,*

$$\boxed{(1 + a)^n \begin{cases} = 1 + na & \text{if } n = 1 \text{ or } a = 0 \\ > 1 + na & \text{if } n > 1 \text{ and } a \neq 0, \end{cases} \qquad \textit{\textbf{Bernoulli's inequality.}}}$$

PROOF. If $a = 0$, then Bernoulli's inequality certainly holds (both sides equal 1), so we'll assume that $a \neq 0$. If $n = 1$, then Bernoulli's inequality is just $1 + a = 1 + a$, which certainly holds. Suppose that Bernoulli's inequality holds for a number n. Then $(1 + a)^n \geq (1 + na)$ (where if $n = 1$, this is an equality and if $n > 1$, this is a strict inequality). Multiplying Bernoulli's inequality by $1 + a > 0$, we obtain

$$(1 + a)^{n+1} \geq (1 + a)(1 + na) = 1 + na + a + na^2.$$

Since $n\,a^2$ is positive, the expression on the right is greater than

$$1 + na + a = 1 + (n + 1)a.$$

Combining this equation with the previous inequality proves Bernoulli's inequality for $n + 1$. By induction, Bernoulli's inequality holds for every $n \in \mathbb{N}$. □

If n is a natural number, recall that the symbol $n!$ (read "n factorial") represents the product of the first n natural numbers. Thus,

$$n! := 1 \cdot 2 \cdot 3 \cdots (n - 1) \cdot n.$$

It is convenient to define $0! = 1$ so that certain formulas continue to hold for $n = 0$. Thus, $n!$ is defined for all nonnegative integers n, that is, $n = 0, 1, 2, \ldots$. Given nonnegative integers n and k with $k \leq n$, we define the **binomial coefficient** $\binom{n}{k}$ by

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

For example, for any nonnegative integer n,

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1 \quad \text{and} \quad \binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!} = 1.$$

Problem 11 contains a generalization of the following important theorem.

THEOREM 2.8 (**Binomial theorem**). *For any complex numbers a and b, and $n \in \mathbb{N}$, we have*

$$\boxed{(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}, \qquad \textbf{\textit{binomial formula}}.}$$

PROOF. If $n = 1$, the right-hand side of the binomial formula reads

$$\sum_{k=0}^{1} \binom{1}{k} a^k\, b^{1-k} = \binom{1}{0} a^0\, b^1 + \binom{1}{1} a^1\, b^0 = b + a,$$

so the binomial formula holds for $n = 1$. Suppose that the binomial formula holds for a natural number n. Then multiplying the formula by $a + b$, we get

$$\begin{aligned}
(a + b)^{n+1} &= (a + b) \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{k+1}\, b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n+1-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{k+1}\, b^{n-k} + a^0\, b^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^k\, b^{n+1-k}. \qquad (2.5)
\end{aligned}$$

Observe that the first term on the right can be rewritten as

$$\begin{aligned}\sum_{k=0}^{n}\binom{n}{k}a^{k+1}\,b^{n-k} &= \binom{n}{0}a^1\,b^n+\binom{n}{1}a^2\,b^{n-1}+\cdots+\binom{n}{n-1}a^n\,b^1+\binom{n}{n}a^{n+1}\,b^0\\ &=\sum_{k=1}^{n}\binom{n}{k-1}a^k\,b^{n+1-k}+a^{n+1}\,b^0. \qquad (2.6)\end{aligned}$$

Also observe that

$$\begin{aligned}\binom{n}{k-1}+\binom{n}{k} &= \frac{n!}{(k-1)!\,(n-k+1)!}+\frac{n!}{k!\,(n-k)!}\\ &=\frac{n!\,k}{k!\,(n-k+1)!}+\frac{n!\,(n-k+1)}{k!\,(n-k+1)!}\\ &=\frac{n!\,(n+1)}{k!\,(n+1-k)!}=\binom{n+1}{k}. \qquad (2.7)\end{aligned}$$

Now replacing (2.6) into (2.5) and using (2.7), we obtain

$$\begin{aligned}(a+b)^{n+1} &= \sum_{k=1}^{n}\binom{n}{k-1}a^k\,b^{n+1-k}+a^{n+1}\,b^0+a^0\,b^{n+1}+\sum_{k=1}^{n}\binom{n}{k}a^k\,b^{n+1-k}\\ &=\sum_{k=1}^{n}\left[\binom{n}{k-1}+\binom{n}{k}\right]a^k\,b^{n+1-k}+a^{n+1}\,b^0+a^0\,b^{n+1}\\ &=a^0\,b^{n+1}+\sum_{k=1}^{n}\binom{n+1}{k}a^k\,b^{n+1-k}+a^{n+1}\,b^0\\ &=\sum_{k=0}^{n+1}\binom{n+1}{k}a^k\,b^{n+1-k}.\end{aligned}$$

Thus, the binomial formula holds for the natural number $n+1$. □

2.2.5. Strong form of induction. Sometimes it is necessary to use the following stronger form of induction. For each natural number n, let P_n be a statement. Suppose that (1) P_1 is true and (2) if n is a natural number and if each statement P_m is true for every $m \le n$, then the statement P_{n+1} is also true. Then every single statement $P_1, P_2, P_3, \ldots$ is true. To see this, let M be all the natural numbers such that P_n is *not* true. We shall prove that M must be empty, which shows that P_n is true for every n. Indeed, suppose that M is not empty. Then by well-ordering, M contains a least element, say n. Since P_1 is true, M does not contain 1, so $n > 1$. Since $1, 2, \ldots, n-1$ are not in M (because n is the least element of M), the statements $P_1, P_2, \ldots, P_{n-1}$ must be true. Hence, by Property (2) of the statements, P_n must also be true. This shows that M does not contain n, which contradicts the assumption that n is in M. Thus, M must be empty. Problems 6, 9, and 10 contain exercises where strong induction is useful.

As already stated, in order to illustrate nontrivial induction examples, in the exercises, we assume basic familiarity with integers, real, and complex numbers.

Exercises 2.2.

1. Consider the statement $1+2+3+\cdots+n=(2n+1)^2/8$. Prove that P_n implies P_{n+1}. What is wrong here?

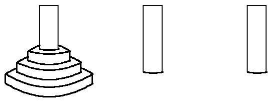


FIGURE 2.3. The towers of Hanoi.

2. Using induction prove that for any complex numbers a and b and for any natural numbers m and n, we have
$$(ab)^n = a^n \cdot b^n \quad , \quad (a^m)^n = a^{mn}.$$
If a and b are nonzero, prove that these equations hold even if $m = 0$ or $n = 0$.
3. Prove the following (some of them quite pretty) formulas/statements via induction:
(a)
$$\frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$
(b)
$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$
(c)
$$1^3 + 2^3 + \cdots + n^3 = (1 + 2 + \cdots + n)^2 = \Big(\frac{n(n+1)}{2}\Big)^2.$$
(d)
$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots + \frac{n}{2^n} = 2 - \frac{n+2}{2^n}.$$
(e) For $a \neq 1$,
$$(1+a)(1+a^2)(1+a^4)\cdots(1+a^{2^n}) = \frac{1-a^{2^{n+1}}}{1-a}.$$
(f) $n^3 - n$ is always divisible by 3.
(g) Every natural number n is either even or odd. Here, n is even means that $n = 2m$ for some $m \in \mathbb{N}$ and n odd means that $n = 1$ or $n = 2m + 1$ for some $m \in \mathbb{N}$.
(h) $n < 2^n$ for all $n \in \mathbb{N}$. (Can you also prove this using Bernoulli's inequality?)
(i) Using the identity (2.7), called **Pascal's rule**, prove that $\binom{n}{k}$ is a natural number for all $n, k \in \mathbb{N}$ with $1 \leq k \leq n$. (P_n is the statement "$\binom{n}{k} \in \mathbb{N}$ for all $1 \leq k \leq n$.")
4. In this problem we prove some nifty binomial formulas. Prove that
$$(a)\ \sum_{k=0}^{n} \binom{n}{k} = 2^n, \qquad (b)\ \sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0,$$
$$(c)\ \sum_{k\ \mathrm{odd}} \binom{n}{k} = 2^{n-1}, \qquad (d)\ \sum_{k\ \mathrm{even}} \binom{n}{k} = 2^{n-1},$$
where the sums in (c) and (d) are over $k = 1, 3, \ldots$ and $k = 0, 2, \ldots$, respectively.
5. (**Towers of Hanoi**) Induction can be used to analyze games! (See Problem 6 in the next section for the game of Nim.) For instance, the *towers of Hanoi* starts with three pegs and n disks of different sizes placed on one peg, with the biggest disk on the bottom and with the sizes decreasing to the top as shown in Figure 2.3. A move is made by taking the top disk off a stack and putting it on another peg so that there is no smaller disk below it. The object of the game is to transfer all the disks to another peg. Prove that the puzzle can be solved in $2^n - 1$ moves. (In fact, you cannot solve the puzzle in less than $2^n - 1$ moves, but the proof of this is another story.)

6. (**The coin game**) Two people have n coins each and they put them on a table, in separate piles, then they take turns removing their own coins; they may take as many as they wish, but they must take at least one. The person removing the last coin(s) wins. Using strong induction, prove that the second person has a "full-proof winning strategy." More explicitly, prove that for each $n \in \mathbb{N}$, there is a strategy such that the second person will win the game with n coins if he follows the strategy.
7. We now prove the **arithmetic-geometric mean inequality** (AGMI): For any nonnegative (that is, ≥ 0) real numbers $a_1, \ldots, a_n$, we have

$$(a_1 \cdots a_n)^{1/n} \leq \frac{a_1 + \cdots + a_n}{n} \quad \text{or equivalently,} \quad a_1 \cdots a_n \leq \Big(\frac{a_1 + \cdots + a_n}{n}\Big)^n.$$

The product $(a_1 \cdots a_n)^{1/n}$ is the **geometric mean** and the sum $\frac{a_1+\cdots+a_n}{n}$ is the **arithmetic mean**, respectively, of the numbers $a_1, \ldots, a_n$.
   (i) Show that $\sqrt{a_1 a_2} \leq \frac{a_1+a_2}{2}$. Suggestion: Expand $(\sqrt{a_1} - \sqrt{a_2})^2 \geq 0$.
   (ii) By induction show the AGMI holds for 2^n terms for every natural number n.
   (iii) Now prove the AGMI for n terms where n is not necessarily a power of 2. (Do *not* use induction.) Suggestion: Let $a = (a_1 + \cdots + a_n)/n$. By Problem 3h, we know that $2^n - n$ is a natural number. Apply the AGMI to the 2^n terms $a_1, \ldots, a_n, a, a, \ldots, a$ where there are $2^n - n$ occurrences of a in this list, to derive the AGMI in general.
8. Here's Newman's proof [**157**] of the AGMI. The AGMI holds for one term, so assume it holds for n terms; we shall prove that the AGMI holds for $n + 1$ terms.
   (a) Prove that if the AGMI holds for all $n+1$ nonnegative numbers $a_1, \ldots, a_{n+1}$ such that $a_1 \cdots a_{n+1} = 1$, then the AGMI holds for any $n + 1$ nonnegative numbers.
   (b) By (a), we just have to verify that the AGMI holds when $a_1 \cdots a_{n+1} = 1$. Using the induction hypothesis, prove that $a_1 + \cdots + a_n + a_{n+1} \geq n(a_{n+1})^{-1/n} + a_{n+1}$.
   (c) Prove that for any $x > 0$, we have $nx^{-1/n} + x \geq n + 1$. Suggestion: Replace n by $n + 1$ and $a = x^{1/n} - 1 > -1$ in Bernoulli's inequality. Now prove that $a_1 + \cdots + a_{n+1} \geq n + 1$, which is the AGMI for $n + 1$ terms.
9. (**Fibonacci sequence**) Define $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. Using strong induction, prove that for every natural number n,

$$F_n = \frac{1}{\sqrt{5}}\Big[\Phi^n - (-\Phi)^{-n}\Big], \qquad \text{where } \Phi = \frac{1+\sqrt{5}}{2} \text{ is called the } \textbf{golden ratio}.$$

Suggestion: Note that $\Phi^2 = \Phi + 1$ and hence $-\Phi^{-1} = 1 - \Phi = (1 - \sqrt{5})/2$.
10. (**Pascal's method**) Using a method due to Pascal, we generalize our formula (2.2) for the sum of the first n integers to sums of powers. See [**18**] for more on Pascal's method. For any natural number k, put $\sigma_k(n) := 1^k + 2^k + \cdots + n^k$ and set $\sigma_0(n) := n$.
   (i) Prove that

$$(n+1)^{k+1} - 1 = \sum_{\ell=0}^{k} \binom{k+1}{\ell} \sigma_\ell(n).$$

   Suggestion: The left-hand side can be written as $\sum_{m=1}^{n} \big((m+1)^{k+1} - m^{k+1}\big)$. Use the binomial theorem on $(m + 1)^{k+1}$.
   (ii) Using the strong form of induction, prove that for each natural number k,

$$\sigma_k(n) = \frac{1}{k+1} n^{k+1} + a_{kk} n^k + \cdots + a_{k2} n^2 + a_{k1} n \quad (\textbf{Pascal's formula}),$$

   for some coefficients $a_{k1}, \ldots, a_{kk} \in \mathbb{Q}$.
   (iii) (Cf. [**124**]) Using the fact that $\sigma_3(n) = \frac{1}{4}n^4 + a_{33}n^3 + a_{32}n^2 + a_{31}n$, find the coefficients a_{31}, a_{32}, a_{33}. Suggestion: Consider the difference $\sigma_3(n) - \sigma_3(n - 1)$. Can you find the coefficients in the sum for $\sigma_4(n)$?
11. (**The multinomial theorem**) A **multi-index** is an n-tuple of nonnegative integers and are usually denoted by Greek letters, for instance $\alpha = (\alpha_1, \ldots, \alpha_n)$ where each

α_k is one of $0, 1, 2, \ldots$. We define $|\alpha| := \alpha_1 + \cdots + \alpha_n$ and $\alpha! := \alpha_1! \cdot \alpha_2! \cdots \alpha_n!$. By induction on n, prove that for any complex numbers $a_1, \ldots, a_n$ and natural number k, we have

$$(a_1 + \cdots + a_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} a_1^{\alpha_1} \cdots a_n^{\alpha_n}.$$

Suggestion: For the induction step, write $a_1 + \cdots + a_{n+1} = a + a_{n+1}$ where $a = a_1 + \cdots + a_n$ and use the binomial formula on $(a + a_{n+1})^k$.

2.3. The integers

Have you ever wondered what it would be like in a world where the temperature was never below zero degrees Celsius? How boring it would be to never see snow! The natural numbers $1, 2, 3, \ldots$ are closed under addition and multiplication, which are essential for counting purposes used in everyday life. However, the natural numbers do not have negatives, which is a problem. In particular, $\mathbb{N}$ is not closed under subtraction. So, e.g. $4 - 7 = -3$, which is not a natural number, therefore the equation

$$x + 7 = 4$$

does not have any solution x in the natural numbers. We can either accept that such an equation does not have solutions[4] or we can describe a number system where such an equation does have solutions. We shall go the latter route and in this section we study the integers or whole numbers, which are closed under subtraction and have negatives.

2.3.1. Axioms for integer numbers.

Incorporating zero and the negatives of the natural numbers,

$$0, -1, -2, -3, -4, \ldots,$$

to the natural numbers forms the **integers** or **whole numbers**:

$$\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots.$$

The set of integers is denoted by $\mathbb{Z}$. The natural numbers are also referred to as the **positive integers**, their negatives the **negative integers**, and the numbers $0, 1, 2, \ldots$, the natural numbers plus zero, are called the **nonnegative integers**, and finally, $0, -1, -2, \ldots$ are the **nonpositive integers**. The following arithmetic properties of addition and multiplication of integers, like for natural numbers, are familiar (in the following a, b denote arbitrary integers):

Addition satisfies

(A1) $a + b = b + a$; (commutative law)

(A2) $(a + b) + c = a + (b + c)$; (associative law)

(A3) there is an integer denoted by 0 “zero” such that

$$a + 0 = a = 0 + a; \quad \text{(existence of additive identity)}$$

(A4) for each integer a there is an integer denoted by the symbol $-a$ such that[5]

$$a + (-a) = 0 \quad \text{and} \quad (-a) + a = 0. \quad \text{(existence of additive inverse)}$$

[4] *The imaginary expression $\sqrt{(-a)}$ and the negative expression $-b$, have this resemblance, that either of them occurring as the solution of a problem indicates some inconsistency or absurdity. As far as real meaning is concerned, both are imaginary, since $0 - a$ is as inconceivable as $\sqrt{(-a)}$. Augustus De Morgan (1806–1871).*

[5] At this moment, there could possibly be another integer, say $b \neq -a$, such that $a + b = 0$, but in Theorem 2.9 we prove that if such a b exists, then $b = -a$; so additive inverses are unique.

Multiplication satisfies

(M1) $a \cdot b = b \cdot a$; (commutative law)
(M2) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$; (associative law)
(M3) there is an integer denoted by 1 "one" such that

$$1 \cdot a = a = a \cdot 1. \quad \text{(existence of multiplicative identity)}$$

As with the natural numbers, the $\cdot$ is sometimes dropped and the associative laws imply that expressions involving integers such as $a+b+c$ or abc make sense without using parentheses. Addition and multiplication are related by

(D) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$. (distributive law)

Of these arithmetic properties, the only additional properties listed that were not listed for natural numbers are (**A3**) and (**A4**). As usual, we denote

$$a + (-b) = (-b) + a \quad \text{by} \quad b - a,$$

so that subtraction is, by definition, really just "adding negatives". A set together with operations of addition and multiplication that satisfy properties (**A1**) – (**A4**), (**M2**), (**M3**), and (**D**) is called a **ring**; essentially a ring is just a set of objects closed under addition, multiplication, and subtraction. If in addition, this multiplication satisfies (M1), then the set is called a **commutative ring**.

The natural numbers or positive integers, which we denote by $\mathbb{N}$ or by $\mathbb{Z}^+$, is closed under addition, multiplication, and has the following property: Given any integer a, exactly one of the following "positivity" properties hold:

(P) a is a positive integer, $a = 0$, or $-a$ is a positive integer.

Stated another way, property (**P**) means that $\mathbb{Z}$ is a union of disjoint sets,

$$\mathbb{Z} = \mathbb{Z}^+ \cup \{0\} \cup -\mathbb{Z}^+,$$

where $-\mathbb{Z}^+$ consists of all integers a such that $-a \in \mathbb{Z}^+$.

Everyday experience convinces us that the integers satisfy properties (**A**), (**M**), (**D**), and (**P**); however, as with the natural numbers, we will *assume* the existence of a set $\mathbb{Z}$ satisfying properties (**A**), (**M**), (**D**), and (**P**) such that $\mathbb{Z}^+ = \mathbb{N}$, the natural numbers. From the properties listed above, we shall derive some well-known properties of the integers memorized since grade school.

2.3.2. Proofs of well-known high school rules. Since the integers satisfy the same arithmetic properties as the natural numbers, the same proofs as in Section 2.1, prove that the distributive law (**D**) holds from the right and the FOIL law holds. Also, the cancellation theorem 2.3 holds: Given any integers a, b, c,

$$a = b \text{ if and only if } a + c = b + c\ .$$

However, now this statement is easily proved using the fact that the integers have additive inverses. We only prove the "if" part: If $a + c = b + c$, then adding $-c$ to both sides of this equation we obtain

$$(a+c)+(-c) = (b+c)+(-c) \implies a+(c+(-c)) = b+(c+(-c)) \implies a+0 = b+0,$$

or $a = b$. Comparing this proof with that of Theorem 2.3 shows the usefulness of having additive inverses.

We now show that we can always solve equations such as the one given at the beginning of this section. Moreover, we prove that there is only one additive identity.

Theorem 2.9 (**Uniqueness of additive identities and inverses**). *For $a, b, x \in \mathbb{Z}$,*

(1) The equation

$$x + a = b \quad \textit{holds if and only if} \quad x = b - a.$$

In particular, the only x that satisfies the equation $x + a = a$ is $x = 0$. Thus, there is only one additive identity.

(2) The only x that satisfies the equation

$$x + a = 0.$$

is $x = -a$. Thus, each integer has only one additive inverse.

(3) Finally, $0 \cdot a = 0$. (zero × anything is zero)

Proof. If $x = b - a$, then

$$(b - a) + a = (b + (-a)) + a = b + ((-a) + a) = b + 0 = b,$$

so the integer $x = b - a$ solves the equation $x + a = b$. Conversely, if x satisfies $x + a = b$, then

$$x + a = b = b + 0 = b + ((-a) + a) = (b + (-a)) + a,$$

so by cancellation (adding $-a$ to both sides), $x = b + (-a) = b - a$. This proves *(1)* and taking $b = 0$ in *(1)* implies *(2)*.

Since $0 = 0 + 0$, we have

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a.$$

Cancelling $0 \cdot a$ (that is, adding $-(0 \cdot a)$ to both sides) proves *(3)*. □

By commutativity, $a + x = b$ if and only if $x = b - a$. Similarly, $a + x = a$ if and only if $x = -a$ and $a + x = 0$ if and only if $x = -a$. We now prove the very familiar "rules of sign" memorized since grade school.

Theorem 2.10 (**Rules of sign**). *The following "rules of signs" hold:*

(1) $-(-a) = a$.
(2) $a \cdot (-1) = -a = (-1) \cdot a$.
(3) $(-1) \cdot (-1) = 1$.
(4) $(-a) \cdot (-b) = ab$.
(5) $(-a) \cdot b = -(ab) = a \cdot (-b)$.
(6) $-(a + b) = (-a) + (-b)$. *In particular,* $-(a - b) = b - a$.

Proof. We prove *(1)*–*(3)* and leave *(4)*–*(6)* for you in Problem 1.

To prove *(1)*, note that since $a + (-a) = 0$, by uniqueness of additive inverses proved in the previous theorem, the additive inverse of $-a$ is a, that is, $-(-a) = a$.

To prove *(2)*, observe that

$$a + a \cdot (-1) = a \cdot 1 + a \cdot (-1) = a \cdot (1 + (-1)) = a \cdot 0 = 0,$$

so by uniqueness of additive inverses, we have $-a = a \cdot (-1)$. By commutativity, $-a = (-1) \cdot a$ also holds.

By *(1)*, *(2)*, we get *(3)*: $(-1) \cdot (-1) = -(-1) = 1$. □

Everyone knows that $-0 = 0$. This fact follows from as an easy application of *(2)*: $-0 = 0 \cdot (-1) = 0$, since zero times anything is zero.

Using the positivity assumption (**P**), we can order the integers in much the same way as the natural numbers are ordered. Given any integers a and b exactly one of the following holds:

(**O1**) $a = b$, that is, $b - a = 0$;
(**O2**) $a < b$, which means that $b - a$ is a positive integer;
(**O3**) $b < a$, which means that $-(b - a)$ is a positive integer.

By our previous theorem, we have $-(b - a) = a - b$, so (**O3**) is just that $a - b$ is a natural number.

Just as for natural numbers, we can define $\leq$, $>$, and $\geq$. For example, $0 < b$, or $b > 0$, means that $b - 0 = b$ is a positive integer. Thus, an integer b is greater than 0 is synonymous with b is a positive integer. (Of course, this agrees with our English usage of $b > 0$ to mean b is positive!) Similarly, $b < 0$ means that $0 - b = -b$ is a positive integer. As with the natural numbers, we have the transitive law: If $a < b$ and $b < c$, then $a < c$, and we also have the Archimedean ordering of $\mathbb{Z}$: Given any natural number a and integer b there is a natural number n so that $b < a \cdot n$. To see this last property, note that if $b < 0$, then any natural number n works; if $b > 0$, then b is a natural number and the Archimedean ordering of $\mathbb{N}$ applies to show the existence of n. We now prove some of the familiar inequality rules.

Theorem 2.11 (**Inequality rules**). *The following inequality rules hold:*

(1) If $a < b$ and $c \leq d$, then $a + c < b + d$.
(2) If $a < b$ and $c > 0$, then $a \cdot c < b \cdot c$. (positives preserve inequalities)
(3) If $a < b$ and $c < 0$, then $a \cdot c > b \cdot c$. (negatives switch inequalities)
(4) If $a > 0$ and $b > 0$, then $ab > 0$. (positive × positive is positive)
(5) If $a > 0$ and $b < 0$ (or vise-verse), then $ab < 0$. (positive × negative is negative)
(6) If $a < 0$ and $b < 0$, then $ab > 0$. (negative × negative is positive)

Proof. We prove *(1)–(3)* and leave *(4)–(6)* for you in Problem 1.

To prove *(1)*, we use associativity and commutativity to write

$$(b + d) - (a + c) = (b - a) + (d - c).$$

Since $a < b$, by definition of less than, $b - a$ is a natural number and since $c \leq d$, $d - c$ is either zero (if $c = d$) or a natural number. Hence, $(b - a) + (d - c)$ is either adding two natural numbers or a natural number and zero; in either case, the result is a natural number. Thus, $a + c < b + d$.

If $a < b$ and $c > 0$, then

$$b \cdot c - a \cdot c = (b - a) \cdot c.$$

c is a natural number and since $a < b$, the integer $b - a$ is a natural number, so their product $(b - a) \cdot c$ is also a natural number. Thus, $a \cdot c < b \cdot c$.

If $a < b$ and $c < 0$, then by our rules of sign,

$$a \cdot c - b \cdot c = (a - b) \cdot c = -(a - b) \cdot (-c) = (b - a) \cdot (-c).$$

Since $c < 0$, the integer $-c$ is a natural number and since $a < b$, the integer $b - a$ is a natural number, so their product $(b - a) \cdot (-c)$ is also a natural number. Thus, $a \cdot c > b \cdot c$. □

We now prove that zero and one have the familiar properties that you know.

THEOREM 2.12 (**Properties of zero and one**). *Zero and one satisfy*

(1) If $a \cdot b = 0$, then $a = 0$ or $b = 0$.

(2) If $a \cdot b = a$ where $a \neq 0$, then $b = 1$; that is, 1 is the only multiplicative identity.

PROOF. We give two proofs of *(1)*. Although **Proof I** is acceptable, **Proof II** is much preferred because **Proof I** boils down to a contrapositive statement anyways, which **Proof II** goes to directly.

Proof I: Assume that $ab = 0$. We shall prove that $a = 0$ or $b = 0$. Now either $a = 0$ or $a \neq 0$. If $a = 0$, then we are done, so assume that $a \neq 0$. We need to prove that $b = 0$. Well, either $b = 0$ or $b \neq 0$. However, it cannot be true that $b \neq 0$, for according to the properties *(4)*, *(5)*, and *(6)* of our rules for inequalities,

$$\text{if } a \neq 0 \text{ and } b \neq 0, \text{ then } a \cdot b \neq 0. \tag{2.8}$$

But $ab = 0$, so $b \neq 0$ cannot be true. This contradiction shows that $b = 0$.

Proof II: Our second proof of *(1)* is a **proof by contraposition**, which is essentially what we did in **Proof I** without stating it! Recall that, already explained in the proof of Theorem 2.3, the technique of a proof by contraposition is that in order to prove the statement "if $a \cdot b = 0$, then $a = 0$ or $b = 0$," we can instead try to prove the contrapositive statement:

$$\text{if } a \neq 0 \text{ and } b \neq 0, \text{ then } a \cdot b \neq 0.$$

However, as explained above (2.8), the truth of this statement follows from our inequality rules. This gives another (better) proof of *(1)*.

To prove *(2)*, assume that $a \cdot b = a$ where $a \neq 0$. Then,

$$0 = a - a = a \cdot b - a \cdot 1 = a \cdot (b - 1).$$

By *(1)*, either $a = 0$ or $b - 1 = 0$. We are given that $a \neq 0$, so we must have $b - 1 = 0$, or adding 1 to both sides, $b = 1$. □

Property *(1)* of this theorem is the basis for solving quadratic equations in high school. For example, let us solve $x^2 - x - 6 = 0$. We first "factor"; that is, observe that

$$(x - 3)(x + 2) = x^2 - x - 6 = 0.$$

By property *(1)*, we know that $x - 3 = 0$ or $x + 2 = 0$. Thus, $x = 3$ or $x = -2$.

2.3.3. Absolute value. Given any integer a, we know that either $a = 0$, a is a positive integer, or $-a$ is a positive integer. The **absolute value** of the integer a is denoted by $|a|$ and is defined to be the "nonnegative part of a":

$$\boxed{|a| := \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{if } a < 0. \end{cases}}$$

Thus, for instance, $|5| = 5$, while $|-2| = -(-2) = 2$. In the following theorem, we prove some (what should be) familiar rules of absolute value. To prove statements about absolute values, it's convenient to **prove by cases**.

THEOREM 2.13 (**Absolute value rules**). *For $a, b, x \in \mathbb{Z}$,*

(1) $|a| = 0$ if and only if $a = 0$.

(2) $|ab| = |a|\,|b|$.

(3) $|a| = |-a|$.

(4) For $x \geq 0$, $|a| \leq x$ if and only if $-x \leq a \leq x$.

(5) $-|a| \leq a \leq |a|$.
(6) $|a+b| \leq |a| + |b|$. ***(triangle inequality)***

Proof. If $a = 0$, then by definition, $|0| = 0$. Conversely, suppose that $|a| = 0$. We have two cases: either $a \geq 0$ or $a < 0$. If $a \geq 0$, then $0 = |a| = a$, so $a = 0$. If $a < 0$, then $0 = |a| = -a$, so $a = 0$ in this case as well. This proves *(1)*.

To prove *(2)*, we consider four cases: $a \geq 0$ and $b \geq 0$, $a < 0$ and $b \geq 0$, $a \geq 0$ and $b < 0$, and lastly, $a < 0$ and $b < 0$. If $a \geq 0$ and $b \geq 0$, then $ab \geq 0$, so $|ab| = ab = |a| \cdot |b|$. If $a < 0$ and $b \geq 0$, then $ab \leq 0$, so $|ab| = -ab = (-a) \cdot b = |a| \cdot |b|$. The case that $a \geq 0$ and $b < 0$ is handled similarly. Lastly, if $a < 0$ and $b < 0$, then $ab > 0$, so $|ab| = ab = (-a) \cdot (-b) = |a| \cdot |b|$.

By *(2)*, we have $|-a| = |(-1) \cdot a| = |-1| \cdot |a| = 1 \cdot |a| = |a|$, which proves *(3)*.

To prove *(4)* we again go to two cases: $a \geq 0$, $a < 0$. In the first case, if $a \geq 0$, then $-x \leq a \leq x$ if and only if $-x \leq |a| \leq x$, which holds if and only if $|a| \leq x$. On the other hand, if $a < 0$, then $-x \leq a \leq x$ if and only if (multiplying through by -1) $-x \leq -a \leq x$ or $-x \leq |a| \leq x$, which holds if and only if $|a| \leq x$. Property *(5)* follows from *(4)* with $x = |a|$.

Finally, we prove the triangle inequality. From *(5)* we have $-|a| \leq a \leq |a|$ and $-|b| \leq b \leq |b|$. Adding these inequalities gives

$$-\big(|a| + |b|\big) \leq a + b \leq |a| + |b|.$$

Applying *(4)* gives the triangle inequality. □

Exercises 2.3.

1. Prove properties *(4)–(6)* in the "Rules of signs" and "Inequality rules" theorems.
2. For integers a, b, prove the inequalities
$$\big|\,|a| - |b|\,\big| \leq |a \pm b| \leq |a| + |b|.$$
3. Let b be an integer. Prove that the only integer a satisfying
$$b - 1 < a < b + 1$$
is the integer $a = b$.
4. Let $n \in \mathbb{N}$. Assume properties of powers from Example 2.2 in Section 2.2.
   (a) Let a, b be *nonnegative* integers. Using a proof by contraposition, prove that if $a^n = b^n$, then $a = b$.
   (b) We now consider the situation when a, b are not necessarily positive. So, let a, b be arbitrary integers. Suppose that $n = 2m$ for some positive integer m. Prove that if $a^n = b^n$, then $a = \pm b$.
   (c) Again let a, b be arbitrary integers. Suppose that $n = 2m - 1$ for some natural number m. Prove the statement if $a^n = b^n$, then $a = b$, using a proof by cases. Here the cases consist of a, b both nonnegative, both negative, and when one is nonnegative and the other negative (in this last case, show that $a^n = b^n$ actually can never hold, so for this last case, the statement is superfluous).
5. In this problem we prove an integer version of induction. Let k be any integer (positive, negative, or zero) and suppose that we are given a list of statements:
$$P_k, P_{k+1}, P_{k+2}, \ldots,$$
and suppose that (1) P_k is true and (2) if n is an integer with $n \geq k$ and the statement P_n happens to be valid, then the statement P_{n+1} is also valid. Then every single statement $P_k, P_{k+1}, P_{k+2}, \ldots$ is true.
6. (**Game of Nim**) Here's a fascinating example using strong induction and proof by cases; see the coin game in Problem 6 of Exercises 2.2 for a related game. Suppose that n stones are thrown on the ground. Two players take turns removing one, two, or

three stones each. The last one to remove a stone loses. Let P_n be the statement that the player starting first has a full-proof winning strategy if n is of the form $n = 4k$, $4k+2$, or $4k+3$ for some $k = 0, 1, 2, \ldots$ and the player starting second has a full-proof winning strategy if $n = 4k+1$ for some $k = 0, 1, 2, \ldots$. In this problem we prove that P_n is true for all $n \in \mathbb{N}$.[6]

(i) Prove that P_1 is true. Assume that $P_1, \ldots, P_n$ hold. To prove that P_{n+1} holds, we prove by cases. The integer $n+1$ can be of four types: $n+1 = 4k$, $n+1 = 4k+1$, $n+1 = 4k+2$, or $n+1 = 4k+3$.

(ii) Case 1: $n+1 = 4k$. The first player can remove one, two, or three stones; in particular, he can remove three stones (leaving $4k-3 = 4(k-1)+1$ stones). Prove that the first person wins.

(iii) Case 2: $n+1 = 4k+1$. Prove that the second player will win regardless if the first person takes one, two, or three stones (leaving $4k$, $4(k-1)+3$, and $4(k-1)+2$ stones, respectively).

(iv) Case 3, Case 4: $n+1 = 4k+2$ or $n+1 = 4k+3$. Prove that the first player has a winning strategy in the cases that $n+1 = 4k+2$ or $n+1 = 4k+3$. Suggestion: Make the first player remove one and two stones, respectively.

2.4. Primes and the fundamental theorem of arithmetic

It is not always true that given any two integers a and b, there is a third integer q (for "quotient") such that

$$b = a\,q.$$

For instance, $2 = 4q$ can never hold for any integer q, nor can $17 = 2q$. This of course is exactly the reason rational numbers are needed! (We shall study rational numbers in Section 2.6.) The existence or nonexistence of such quotients opens up an incredible wealth of topics concerning prime numbers in number theory.[7]

2.4.1. Divisibility.

If a and b are integers, and there is a third integer q such that

$$b = a\,q,$$

then we say that a **divides** b or b **divisible** by a or b is a **multiple** of a, in which case we write $a|b$ and call a a **divisor** or **factor** of b and q the **quotient** (of b divided by a).

Example 2.6. Thus, for example $4|(-16)$ with quotient 4 because $-16 = 4 \cdot (-4)$ and $(-2)|(-6)$ with quotient 3 because $-6 = (-2) \cdot 3$.

We also take the convention that

divisors are by definition nonzero.

To see why, note that

$$0 = 0 \cdot 0 = 0 \cdot 1 = 0 \cdot (-1) = 0 \cdot 2 = 0 \cdot (-2) = \cdots,$$

[6] Here, we are implicitly assuming that any natural number can be written in the form $4k$, $4k+1$, $4k+2$, or $4k+3$; this follows from Theorem 2.15 on the division algorithm, which we assume just for the sake of presenting a cool exercise!

[7] *Mathematicians have tried in vain to this day to discover some order in the sequence of prime numbers, and we have reason to believe that it is a mystery into which the human mind will never penetrate. Leonhard Euler (1707–1783)* [**210**].

so if 0 were allowed to be a divisor, then every integer is a quotient when 0 is divided by itself! However, if $a \neq 0$, then $b = aq$ can have only one quotient, for if in addition $b = aq'$ for some integer q', then

$$aq = aq' \quad \Longrightarrow \quad aq - aq' = 0 \quad \Longrightarrow \quad a(q - q') = 0.$$

Since $a \neq 0$, we must have $q - q' = 0$, or $q = q'$. So the quotient q is unique. Because uniqueness is so important in mathematics we always assume that divisors are nonzero. Thus, comes the high school phrase "You can never divide by 0!" Here are some important properties of division.

Theorem 2.14 (**Divisibility rules**). *The following divisibility rules hold:*

(1) If $a|b$ and b is positive, then $|a| \leq b$.
(2) If $a|b$, then $a|bc$ for any integer c.
(3) If $a|b$ and $b|c$, then $a|c$.
(4) If $a|b$ and $a|c$, then $a|(bx + cy)$ for any integers x and y.

Proof. Assume that $a|b$ and $b > 0$. Since $a|b$, we know that $b = aq$ for some integer q. Assume for the moment that $a > 0$. By our inequality rules (Theorem 2.11), we know that "positive × negative is negative", so q cannot be negative. q also can't be zero because $b \neq 0$. Therefore $q > 0$. By our rules for inequalities,

$$a = a \cdot 1 \leq a \cdot q = b \quad \Longrightarrow \quad |a| \leq b.$$

Assume now that $a < 0$. Then $b = aq = (-a)(-q)$. Since $(-a) > 0$, by our proof for positive divisors that we just did, we have $(-a) \leq b$, that is, $|a| \leq b$.

We now prove *(2)*. If $a|b$, then $b = aq$ for some integer q. Hence,

$$bc = (aq)c = a(qc),$$

so $a|bc$.

To prove *(3)*, suppose that $a|b$ and $b|c$. Then $b = aq$ and $c = bq'$ for some integers q and q'. Hence,

$$c = bq' = (aq)q' = a(qq'),$$

so $a|c$.

Finally, assume that $a|b$ and $a|c$. Then $b = aq$ and $c = aq'$ for integers q and q'. Hence, for any integers x and y,

$$bx + cy = (aq)x + (aq')y = a(qx + q'y),$$

so $a|(bx + cy)$. □

2.4.2. The division algorithm. Although we cannot always divide one integer into another we can always do it up to remainders.

Example 2.7. For example, although 2 does not divide 7, we can write

$$7 = 3 \cdot 2 + 1.$$

Another example is that although -3 does not divide -13, we do have

$$-13 = 5 \cdot (-3) + 2.$$

In general, if a and b are integers and

$$b = qa + r, \quad \text{where } 0 \leq r < |a|,$$

then we call q the **quotient** (of b divided by a) and r the **remainder**. Such numbers always exists as we now prove.

THEOREM 2.15 (**The division algorithm**). *Given any integers a and b with $a \neq 0$, there are unique integers q and r so that*

$$\boxed{b = qa + r \quad \text{with} \quad 0 \le r < |a|.}$$

Moreover, if a and b are both positive, then q is nonnegative. Furthermore, a divides b if and only if $r = 0$.

PROOF. Assume for the moment that $a > 0$. Consider the list of integers

$$\ldots, 1+b-3a,\ 1+b-2a,\ 1+b-a,\ b,\ 1+b+a,\ 1+b+2a,\ 1+b+3a, \ldots \tag{2.9}$$

extending indefinitely in both ways. Notice that since $a > 0$, for any integer n,

$$1 + b + na < 1 + b + (n+1)a,$$

so the integers in the list (2.9) are increasing. Moreover, by the Archimedean ordering of the integers, there is a natural number n so that $-1 - b < an$ or $1+b+an > 0$. In particular, $1+b+ak > 0$ for $k \ge n$. Thus, far enough to the right in the list (2.9), all the integers are positive. Let A be set of all natural numbers appearing in the list (2.9). By the well-ordering principle (Theorem 2.6), this set of natural numbers has a least element, let us call it $1+b+ma$ where m is an integer. This integer satisfies

$$1 + b + (m-1)a < 1 \le 1 + b + ma, \tag{2.10}$$

for if $1 + b + (m-1)a \ge 1$, then $1 + b + (m-1)a$ would be an element of A smaller than $1 + b + ma$. Put $q = -m$ and $r = b + ma = b - qa$. Then $b = qa + r$ by construction, and substituting in q and r into (2.10), we obtain

$$1 + r - a < 1 \le 1 + r.$$

Subtracting 1 from everything, we see that

$$r - a < 0 \le r.$$

Thus, $0 \le r$ and $r - a < 0$ (that is, $r < a$). Thus, we have found integers q and r so that $b = qa + r$ with $0 \le r < a$. Observe from (2.10) that if b is positive, then m can't be positive (for otherwise the left-hand inequality in (2.10) wouldn't hold). Thus, q is nonnegative if both a and b are positive. Assume now that $a < 0$. Then $-a > 0$, so by what we just did, there are integers s and r with $b = s(-a) + r$ with $0 \le r < -a$; that is, $b = qa + r$, where $q = -s$ and $0 \le r < |a|$.

We now prove uniqueness. Assume that we also have $b = q'a + r'$ with $0 \le r' < |a|$. We first prove that $r = r'$. Indeed, suppose that $r \neq r'$, then by symmetry in the primed and unprimed letters, we may presume that $r < r'$. Then $0 < r' - r \le r' < |a|$. Moreover,

$$q'a + r' = qa + r \quad \Longrightarrow \quad (q' - q)a = r' - r.$$

This shows that $a|(r' - r)$ which is impossible since $r' - r$ is smaller than $|a|$ (see property *(1)* of Theorem 2.14). Thus, we must have $r = r'$. Then the equation $(q' - q)a = r' - r$ reads $(q' - q)a = 0$. Since $a \neq 0$, we must have $q' - q = 0$, or $q = q'$. Our proof of uniqueness is thus complete.

Finally, we prove that $a|b$ if and only if $r = 0$. If $a|b$, then $b = ac = ac + 0$ for some integer c. By uniqueness already established, we have $q = c$ and $r = 0$. Conversely, if $r = 0$, then $b = aq$, so $a|b$ by definition of divisibility. □

An integer n is **even** if we can write $n = 2m$ for some integer m, and **odd** if we can write $n = 2m + 1$ for some integer m.

Example 2.8. For instance, $0 = 2 \cdot 0$ so 0 is even, $1 = 2 \cdot 0 + 1$ so 1 is odd, and -1 is odd since $-1 = 2 \cdot (-1) + 1$.

Using the division algorithm we can easily prove that every integer is either even or odd. Indeed, dividing n by 2, the division algorithm implies that $n = 2m + k$ where $0 \le k < 2$, that is, where k is either 0 or 1. This shows that n is either even (if $k = 0$) or odd (if $k = 1$).

An important application of the division algorithm is to the so-called Euclidean algorithm for finding greatest common divisors; see Problem 4.

2.4.3. Prime numbers. Consider the number 12. This number has 6 positive factors or divisors, 1, 2, 3, 4, 6, and 12. The number 21 has 4 positive factors, 1, 3, 7, and 21. The number 1 has only one positive divisor, 1. However, as the reader can check, 17 has exactly two positive factors, 1 and 17. Similarly, 5 has exactly two positive factors, 1 an 5. Numbers such as 5 and 17 are given a special name: A natural number that has exactly two positive factors is called a **prime** number.[8] Another way to say this is that a prime number is a natural number with exactly two factors, itself and 1. (Thus, 1 is not prime.) A list of the first ten primes is

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \ldots.$$

A number that is not prime is called a **composite** number. Notice that

$$12 = 2 \times 6 = 2 \times 2 \times 3, \quad 21 = 3 \times 7, \quad 17 = 17, \quad 5 = 5.$$

In each of these circumstances, we have **factored** or expressed as a product, each number into a product of its prime factors. Here, by *convention*, we consider a prime number to be already in its factored form.

Lemma 2.16. *Every natural number other than* 1 *can be factored into primes.*

Proof. We shall prove that for any natural number $m = 1, 2, 3, \ldots$, the number $m + 1$ can be factored into primes; which is to say, any natural number $n = 2, 3, 4, \ldots$ can be factored into primes. We prove this lemma by using strong induction. By our convention, $n = 2$ is already in factored form. Assume our theorem holds for all natural numbers $2, 3, 4, \ldots, n$; we shall prove our theorem holds for the natural number $n + 1$. Now $n + 1$ is either prime or composite. If it is prime, then it is already in factored form. If it is composite, then $n + 1 = pq$ where p and q are natural numbers greater than 1. By Theorem 2.14 both p and q are less than $n + 1$. By induction hypothesis, p and q can be factored into primes. It follows that $n + 1 = pq$ can also be factored into primes. □

One of the first questions that one may ask is how many primes there are. This was answered by Euclid of Alexandria (325 B.C.–265 B.C.): There are infinity many. The following proof is the original due to Euclid and is the classic "proof by contradiction".

Theorem 2.17 (**Euclid's theorem**). *There are infinitely many primes.*

[8] *I hope you will agree that there is no apparent reason why one number is prime and another not. To the contrary, upon looking at these numbers one has the feeling of being in the presence of one of the inexplicable secrets of creation. Don Bernard Zagier* [**253**, p. 8].

Proof. We start with the tentative assumption that the theorem is false. Thus, we assume that there are only finitely many primes. There being only finitely many, we can list them:

$$p_1,\ p_2,\ \ldots, p_n.$$

Consider the number

$$p_1p_2p_3\cdots p_n + 1.$$

This number is either prime or composite. It is greater than all the primes $p_1, \ldots, p_n$, so this number can't equal any $p_1, \ldots, p_n$. We conclude that n must be composite, so

$$p_1p_2p_3\cdots p_n + 1 = ab, \tag{2.11}$$

for some natural numbers a and b. By our lemma, both a and b can be expressed as a product involving $p_1, \ldots, p_n$, which implies that ab also has such an expression. In particular, being a product of some of the $p_1, \ldots, p_n$, the right-hand side of (2.11) is divisible by at least one of the prime numbers $p_1, \ldots, p_n$. However, the left-hand side is certainly not divisible by any such prime because if we divide the left-hand side by any one of the primes $p_1, \ldots, p_n$, we always get the remainder 1! This contradiction shows that our original assumption that the theorem is false must have been incorrect; hence there must be infinitely many primes. □

2.4.4. Fundamental theorem of arithmetic. Consider the integer 120, which can be factored as follows:

$$120 = 2 \times 2 \times 2 \times 3 \times 5.$$

A little verification shows that it is impossible to factor 120 into any primes other than the ones displayed. Of course, the order can be different; e.g.

$$120 = 3 \times 2 \times 2 \times 5 \times 2.$$

It is of fundamental importance in mathematics that any natural number can be factored into a product of primes in only one way, apart from the order.

Theorem 2.18 (**Fundamental theorem of arithmetic**). *Every natural number other than* 1 *can be factored into primes in only one way, except for the order of the factors.*

Proof. For sake of contradiction, let us suppose that there are primes that can be factored in more that one way. By the well-ordering principle, there is a smallest such natural number a. Thus, we can write a as a product of primes in two ways:

$$a = p_1p_2\cdots p_m = q_1q_2\cdots q_n.$$

Note that both m and n are greater than 1, for a single prime number has one prime factorization. We shall obtain a contradiction by showing there is a smaller natural number that has two factorizations. First, we observe that none of the primes p_j on the left equals any of the primes q_k on the right. Indeed, if for example $p_1 = q_1$, then by cancellation, we could divide them out obtaining the natural number

$$p_2p_3\cdots p_m = q_2q_3\cdots q_n.$$

This number is smaller than a and the two sides must represent two distinct prime factorizations, for if these prime factorizations were the same apart from the orderings, then (since $p_1 = q_1$) the factorizations for a would also be the same apart from orderings. Since a is the smallest such number with more than one factorization, we conclude that none of the primes p_j equals a prime q_k. In particular, since $p_1 \neq q_1$, by symmetry we may assume that $p_1 < q_1$. Now consider the natural number

$$
\begin{aligned}
b &= (q_1 - p_1)q_2q_3 \cdots q_n \qquad (2.12)\\
&= q_1q_2 \cdots q_n - p_1q_2 \cdots q_n\\
&= p_1p_2 \cdots p_m - p_1q_2 \cdots q_n \quad (\text{since } p_1 \cdots p_m = a = q_1 \cdots q_m)\\
&= p_1(p_2p_3 \cdots p_m - q_2q_3 \cdots q_n). \qquad (2.13)
\end{aligned}
$$

Since $0 < q_1 - p_1 < q_1$, the number b is less than a, so b can only be factored in one way apart from orderings. Observe that the number $q_1 - p_1$ cannot have p_1 as a factor, for if p_1 divides $q_1 - p_1$, then p_1 also divides $(q_1 - p_1) + p_1 = q_1$, which is impossible because q_1 is prime. Thus, writing $q_1 - p_1$ into its prime factors, none of which is p_1, the expression (2.12) and the fact that $p_1 \neq q_k$ for any k shows that b does not contain the factor p_1 in its factorization. On the other hand, writing $p_2p_3 \cdots p_m - q_2q_3 \cdots q_n$ into its prime factors, the expression (2.13) clearly shows that p_1 is in the prime factorization of b. This contradiction ends the proof. □

Another popular way to prove the fundamental theorem of arithmetic uses the concept of the **greatest common divisor**; see Problem 5 for this proof.

In our first exercise, recall that the notation $n!$ (read "n factorial") for $n \in \mathbb{N}$ denotes the product of the first n integers: $n! := 1 \cdot 2 \cdot 3 \cdots n$.

Exercises 2.4.

1. A natural question is: How sparse are the primes? Prove that there are arbitrarily large gaps in the list of primes in the following sense: Given any positive integer k, there are k consecutive composite integers. Suggestion: Consider the integers
$$(k+1)! + 2,\ (k+1)! + 3, \ldots, (k+1)! + k,\ (k+1)! + k + 1.$$
2. Using the fundamental theorem of arithmetic, prove that if a prime p divides ab, where $a, b \in \mathbb{N}$, then p divides a or p divides b. Is this statement true if p is not prime?
3. Prove Lemma 2.16, that every natural number other than 1 can be factored into primes, using the well-ordering principle instead of induction.
4. (**The Euclidean algorithm**) Let a and b be any two integers, both not zero. Consider the set of all positive integers that divide both a and b. This set is nonempty (it contains 1) and is finite (since integers larger than $|a|$ and $|b|$ cannot divide both a and b). This set therefore has a largest element (Problem 6 in Exercises 2.1), which we denote by (a, b) and call the **greatest common divisor** (GCD) of a and b. In this problem we find the GCD using the Euclidean algorithm.
   (i) Show that $(\pm a, b) = (a, \pm b) = (a, b)$ and $(0, b) = |b|$. Because of these equalities, we henceforth assume that a and b are positive.
   (ii) By the division algorithm we know that there are unique nonnegative integers q_0 and r_0 so that $b = q_0a + r_0$ with $0 \le r_0 < a$. Show that $(a, b) = (a, r_0)$.
   (iii) By successive divisions by remainders, we can write
$$
\begin{gathered}
b = q_0 \cdot a + r_0, \quad a = q_1 \cdot r_0 + r_1, \quad r_0 = q_2 \cdot r_1 + r_2,\\
r_1 = q_3 \cdot r_2 + r_3, \quad \ldots \quad r_{j-1} = q_j \cdot r_j + r_{j+1}, \quad \ldots,
\end{gathered} \qquad (2.14)
$$
   where the process is continued only as far as we don't get a zero remainder. Show that $a > r_0 > r_1 > r_2 > \cdots$ and using this fact, explain why we must eventually get a zero remainder.
   (iv) Let $r_{n+1} = 0$ be the first zero remainder. Show that $r_n = (a, b)$. Thus, the last positive remainder in the sequence (2.14) equals (a, b). This process for finding the GCD is called the **Euclidean algorithm**.
   (v) Using the Euclidean algorithm, find $(77, 187)$ and $(193, 245)$.

5. Working backwards through the equations (2.14) show that for any two integers a, b, we have

$$(a, b) = r_n = k\,a + \ell\,b,$$

for some integers k and ℓ. Using this fact concerning the GCD, we shall give an easy proof of the fundamental theorem of arithmetic.

(i) Prove that if a prime p divides a product ab, then p divides a or p divides b. (Problem 2 does not apply here because in that problem we used the fundamental theorem of arithmetic, but now we are going to prove this fundamental theorem.) Suggestion: Either p divides a or it doesn't; if it does, we're done, if not, then the GCD of p and a is 1. Thus, $1 = (p, a) = k\,a + \ell\,b$, for some integers k, ℓ. Multiply this equation by b and show that p must divide b.

(ii) Using induction prove that if a prime p divides a product $a_1 \cdots a_n$, then p divides some a_i.

(iii) Using (ii), prove that the fundamental theorem of arithmetic.

6. (**Modular arithmetic**) Given $n \in \mathbb{N}$, we say that $x, y \in \mathbb{Z}$ are **congruent modulo** n, written $x \equiv y \pmod{n}$, if $x - y$ is divisible by n. For $a, b, x, y, u, v \in \mathbb{Z}$, prove

(a) $x \equiv y \pmod{n}$, $y \equiv x \pmod{n}$, $x - y \equiv 0 \pmod{n}$ are equivalent statements.

(b) If $x \equiv y \pmod{n}$ and $y \equiv z \pmod{n}$, then $x \equiv z \pmod{n}$.

(c) If $x \equiv y \pmod{n}$ and $u \equiv v \pmod{n}$, then $ax + by \equiv au + bv \pmod{n}$.

(d) If $x \equiv y \pmod{n}$ and $u \equiv v \pmod{n}$, then $xu \equiv yv \pmod{n}$.

(e) Finally, prove that if $x \equiv y \pmod{n}$ and $m|n$ where $m \in \mathbb{N}$, then $x \equiv y \pmod{m}$.

7. (**Fermat's theorem**) We assume the basics of modular arithmetic from Problem 6. In this problem we prove that for any prime p and $x \in \mathbb{Z}$, we have $x^p \equiv x \pmod{p}$. This theorem is due to Pierre de Fermat (1601–1665).

(i) Prove that for any $k \in \mathbb{N}$ with $1 < k < p$, the binomial coefficient (which is an integer, see e.g. Problem 3i in Exercises 2.2) $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ is divisible by p.

(ii) Using (i), prove that for any $x, y \in \mathbb{Z}$, $(x + y)^p \equiv x^p + y^p \pmod{p}$.

(iii) Using (ii) and induction, prove that $x^p \equiv x \pmod{p}$ for all $x \in \mathbb{N}$. Conclude that $x^p \equiv x \pmod{p}$ for all $x \in \mathbb{Z}$.

8. (**Pythagorean triples**) A **Pythagorean triple** consists of three natural numbers (x, y, z) such that $x^2 + y^2 = z^2$. For example, $(3, 4, 5)$ and $(6, 8, 10)$ are such triples. The triple is called **primitive** if x, y, z are **relatively prime**, or **coprime**, which means that x, y, z have no common prime factors. For instance, $(3, 4, 5)$ is primitive while $(6, 8, 10)$ is not. In this problem we prove

$$(x, y, z) \text{ is primitive } \iff \begin{cases} x = 2mn\,,\ y = m^2 - n^2\,,\ z = m^2 + n^2, & \text{or,} \\ x = m^2 - n^2\,,\ y = 2mn\,,\ z = m^2 + n^2, \end{cases}$$

where m, n are coprime, $m > n$, and m, n are of opposite parity; that is, one of m, n is even and the other is odd.

(i) Prove the "$\Longleftarrow$" implication. Henceforth, let (x, y, z) be a primitive triple.

(ii) Prove that x and y cannot both be even.

(iii) Show that x and y cannot both be odd.

(iv) Therefore, one of x, y is even and the other is odd; let us choose x as even and y as odd. (The other way around is handled similarly.) Show that z is odd and conclude that $u = \frac{1}{2}(z + y)$ and $v = \frac{1}{2}(z - y)$ are both natural numbers.

(v) Show that $y = u - v$ and $z = u + v$ and then $x^2 = 4uv$. Conclude that uv is a perfect square (that is, $uv = k^2$ for some $k \in \mathbb{N}$).

(vi) Prove that u and v must be coprime and from this fact and the fact that uv is a perfect square, conclude that u and v each must be a perfect square; say $u = m^2$ and $v = n^2$ for some $m, n \in \mathbb{N}$. Finally, prove the desired result.

9. (**Pythagorean triples, again**) If you like primitive Pythagorean triples, here's another problem: Prove that if m, n are coprime, $m > n$, and m, n are of the *same* parity,

then

$$(x,y,z) \text{ is primitive,} \quad \text{where } x = mn\,,\ y = \frac{m^2-n^2}{2}\,,\ z = \frac{m^2+n^2}{2}.$$

Combined with the previous problem, we see that given coprime natural numbers $m > n$,

$$(x,y,z) \text{ is primitive,} \quad \text{where } \begin{cases} x = 2mn\,,\ y = m^2-n^2\,,\ z = m^2+n^2, & \text{or,} \\ x = mn\,,\ y = \frac{m^2-n^2}{2}\,,\ z = \frac{m^2+n^2}{2}, \end{cases}$$

according as m and n have opposite or the same parity.

10. (**Mersenne primes**) A number of the form $M_n = 2^n - 1$ is called a **Mersenne number**, named after Marin Mersenne (1588–1648). If M_n is prime, it's called a **Mersenne prime**. For instance, when $M_2 = 2^2 - 1 = 3$ is prime, $M_3 = 2^3 - 1 = 7$ is prime, but $M_4 = 2^4 - 1 = 15$ is not prime. However, when $M_5 = 2^5 - 1 = 31$ is prime again. It it not known if there exists infinitely many Mersenne primes. Prove that if M_n is prime, then n is prime. (The converse if false; for instance M_{23}, is composite.) Suggestion: Prove the contrapositive. Also, the polynomial identity $x^k - 1 = (x-1)(x^{k-1} + x^{k-2} + \cdots + x + 1)$ might be helpful.

11. (**Perfect numbers**) A number $n \in \mathbb{N}$ is said to be **perfect** if it is the sum of its proper divisors (divisors excluding itself). For example, $6 = 1+2+3$ and $28 = 1+2+4+7+14$ are perfect. It's not known if there exists any odd perfect numbers! In this problem we prove that perfect numbers are related to Mersenne primes as follows:

$$n \text{ is even and perfect} \iff n = 2^m(2^{m+1}-1) \text{ where } m \in \mathbb{N}\,,\ 2^{m+1}-1 \ \text{ is prime.}$$

For instance, when $m = 1$, $2^{1+1} - 1 = 3$ is prime, so $2^1(2^{1+1} - 1) = 6$ is perfect. Similarly, we get 28 when $m = 2$ and the next perfect number is 496 when $m = 4$. (Note that when $m = 3$, $2^{m+1} - 1 = 15$ is not prime.)

(i) Prove that if $n = 2^m(2^{m+1} - 1)$ where $m \in \mathbb{N}$ and $2^{m+1} - 1$ is prime, then n is perfect. Suggestion: The proper divisors of n are $1, 2, \ldots, 2^m, q, 2q, \ldots, 2^{m-1}q$ where $q = 2^{m+1} - 1$.

(ii) To prove the converse, we proceed systematically as follows. First prove that if $m, n \in \mathbb{N}$, then d is a divisor of $m \cdot n$ if and only if $d = d_1 \cdot d_2$ where d_1 and d_2 are divisors of m and n, respectively. Suggestion: Write $m = p_1^{m_1} \cdots p_k^{m_k}$ and $n = q_1^{n_1} \cdots q_\ell^{n_\ell}$ into prime factors. Observe that a divisor of $m \cdot n$ is just a number of the form $p_1^{i_1} \cdots p_k^{i_k} q_1^{j_1} \cdots q_\ell^{j_\ell}$ where $0 \le i_r \le m_r$ and $0 \le j_r \le n_r$.

(iii) For $n \in \mathbb{N}$, define $\sigma(n)$ as the sum of all the divisors of n (including n itself). Using (ii), prove that if $m, n \in \mathbb{N}$, then $\sigma(m \cdot n) = \sigma(m) \cdot \sigma(n)$.

(iv) Let n be even and perfect and write $n = 2^m q$ where $m \in \mathbb{N}$ and q is odd. By (iii), $\sigma(n) = \sigma(2^m)\sigma(q)$. Working out both sides of $\sigma(n) = \sigma(2^m)\sigma(q)$, prove that

$$\sigma(q) = q + \frac{q}{2^{m+1}-1}. \tag{2.15}$$

Suggestion: Since n is perfect, prove that $\sigma(n) = 2n$ and by definition of σ, prove that $\sigma(2^m) = 2^{m+1} - 1$.

(v) From (2.15) and the fact that $\sigma(q) \in \mathbb{N}$, show that $q = k(2^{m+1} - 1)$ for some $k \in \mathbb{N}$. From (2.15) (that $\sigma(q) = q + k$), prove that $k = 1$. Finally, conclude that $n = 2^m(2^{m+1} - 1)$ where $q = 2^{m+1} - 1$ is prime.

12. In this exercise we show how to factor factorials (cf. [**78**]). Let $n > 1$. Show that the prime factors of $n!$ are exactly those primes less than or equal to n. Explain that to factor $n!$, for each prime p less than n, we need to know the greatest power of p that divides $n!$. We shall prove that the greatest power of p that divides $n!$ is

$$e_p(n) := \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor, \tag{2.16}$$

where $\lfloor n/p^k \rfloor$ is the quotient when n is divided by p^k.

(a) If $p^k > n$, show that $\lfloor n/p^k \rfloor = 0$, so the sum in (2.16) is actually finite.

(b) Show that

$$\left\lfloor \frac{n+1}{p^k} \right\rfloor - \left\lfloor \frac{n}{p^k} \right\rfloor = \begin{cases} 1 & \text{if } p^k \,|\, (n+1) \\ 0 & \text{if } p^k \nmid (n+1). \end{cases}$$

(c) Prove that

$$\sum_{k=1}^{\infty} \left\lfloor \frac{n+1}{p^k} \right\rfloor - \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor = j,$$

where j is the largest integer such that p^j divides $n+1$.

(d) Now prove (2.16) by induction on n. Suggestion: For the induction step, write $(n+1)! = (n+1)\,n!$ and show that $e_p(n+1) = j\,e_p(n)$, where j is the largest integer such that p^j divides $n+1$.

(e) Use (2.16) to find $e_2, e_3, e_5, e_7, e_{11}$ for $n = 12$ and then factor 12! into primes.

2.5. Decimal representations of integers

Since grade school we have represented numbers in "base 10". In this section we explore the use of arbitrary bases.

2.5.1. Decimal representations of integers. We need to carefully make a distinction between integers and the symbols used to represent them.

Example 2.9. In our common day notation, we have

$$2 = 1 + 1, \quad 3 = 2 + 1, \quad 4 = 3 + 1, \ldots \text{etc.}$$

where 1 is our symbol for the multiplicative unit.

Example 2.10. The Romans used the symbol I for the multiplicative unit, and for the other numbers,

$$II = I + I, \quad III = II + I, \quad IV = III + I, \ldots \text{etc;}$$

if you want to be proficient in using Roman numerals, see [**208**].

Example 2.11. We could be creative and make up our own symbols for integers: e.g.

$$\text{i} = 1 + 1, \quad \text{like} = \text{i} + 1, \quad \text{math} = \text{like} + 1, \ldots \text{etc.}$$

As you could imagine, it would be very inconvenient to make up a different symbol for every single number! For this reason, we write numbers with respect to "bases". For instance, undoubtedly because we have ten fingers, the base 10 or **decimal** system, is the most widespread system to make symbols for the integers. In this system, we use the symbols $0, 1, 2, \ldots, 9$ called **digits** for zero and the first nine positive integers, to give a symbol to any integer using the symbol $10 := 9 + 1$ as the "base" with which to express numbers.

Example 2.12. Consider the *symbol* 12. This symbol represents the *number* twelve, which is the number

$$1 \cdot 10 + 2.$$

Example 2.13. The symbol 4321 represents the number a given in words by four thousand, three hundred and twenty-one:

$$a = 4000 + 300 + 20 + 1 = 4 \cdot 10^3 + 3 \cdot 10^2 + 2 \cdot 10 + 1.$$

Note that the digits $1, 2, 3, 4$ in the symbol 4321 are exactly the remainders produced after successive divisions of a and its quotients by 10. For example,

$$a = 432 \cdot 10 + 1 \quad \text{(remainder 1)}$$

Now divide the quotient 432 by 10:

$$432 = 43 \cdot 10 + 2 \quad \text{(remainder 2)}.$$

Continuing dividing the quotients by 10, we get

$$43 = 4 \cdot 10 + 3, \quad \text{(remainder 3)}, \qquad \text{and finally,} \quad 4 = 0 \cdot 10 + 4, \quad \text{(remainder 4)}.$$

We shall use this technique of successive divisions in the proof of Theorem 2.19 below. In general, the *symbol* $a = a_n a_{n-1} \cdots a_1 a_0$ represents the *number*

$$a = a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \cdots + a_1 \cdot 10 + a_0 \quad \text{(in base 10)}.$$

As with our previous example, the digits $a_0, a_1, \ldots, a_n$ are exactly the remainders produced after successive divisions of a and the resulting quotients by 10.

2.5.2. Other common bases. We now consider other bases; for instance, the base 7 or **septimal** system. Here, we use the symbols $0, 1, 2, 3, 4, 5, 6, 7$ to represent zero and the first seven natural numbers and the numbers $0, 1, \ldots, 6$ are the **digits** in base 7. Then we write an integer a as $a_n a_{n-1} \cdots a_1 a_0$ in base 7 if

$$a = a_n \cdot 7^n + a_{n-1} \cdot 7^{n-1} + \cdots + a_1 \cdot 7 + a_0.$$

Example 2.14. For instance, the number with symbol 10 in base 7 is really the number 7 itself, since

$$10 \text{ (base 7)} = 1 \cdot 7 + 0.$$

Example 2.15. The number one hundred one has the symbol 203 in the septimal system because

$$203 \text{ (base 7)} = 2 \cdot 7^2 + 0 \cdot 7 + 3,$$

and in our familiar base 10 or decimal notation, the number on the right is just $2 \cdot 49 + 3 = 98 + 3 = 101$.

The base of choice for computers is base 2 or the **binary** or **dyadic** system. In this case, we write numbers using only the digits 0 and 1. Thus, an integer a is written as $a_n a_{n-1} \cdots a_1 a_0$ in base 2 if

$$a = a_n \cdot 2^n + a_{n-1} \cdot 2^{n-1} + \cdots + a_1 \cdot 2 + a_0.$$

Example 2.16. For instance, the symbol 10101 in the binary system represents the number

$$10101 \text{ (base 2)} = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1.$$

In familiar base 10 or decimal notation, the number on the right is $16+4+2+1 = 21$.

Example 2.17. The symbol 10 in base 2 is really the number 2 itself, since

$$10 \text{ (base 2)} = 1 \cdot 2 + 0.$$

Not only are binary numbers useful for computing, they can also help you be a champion in the *Game of Nim*; see [**202**]. (See also Problem 6 in Exercises 2.3.) Another common base is base 3, which is known as the **tertiary** system.

We remark that one can develop addition and multiplication tables in the septimal and binary systems (indeed, with respect to any base); see for instance [**57**, p. 7]. Once a base is fixed, we shall not make a distinction between a number and its representation in the chosen base. In particular, throughout this book we always use base 10 and write numbers with respect to this base unless stated otherwise.

2.5.3. Arbitrary base expansions of integers. We now show that a number can be written with respect to any base. Fix a natural number $b > 1$, called a **base**. Let a be a natural number and suppose that it can be written as a sum of the form

$$\boxed{a = a_n \cdot b^n + a_{n-1} \cdot b^{n-1} + \cdots + a_1 \cdot b + a_0,}$$

where $0 \leq a_k < b$ and $a_n \neq 0$. Then the symbol $a_n a_{n-1} \cdots a_1 a_0$ is called the **b-adic representation** of a. A couple questions arise: First, does every natural number have such a representation and second, if a representation exists, is it unique? The answer to both questions is yes.

In the following proof, we shall use the following useful "telescoping" sum several times:

$$\sum_{k=0}^{n}(b-1)\,b^k = \sum_{k=0}^{n}(b^{k+1} - b^k) = b^1 + \cdots + b^n + b^{n+1} - (1 + b^1 + \cdots + b^n) = b^{n+1} - 1.$$

THEOREM 2.19. *Every natural number has a unique b-adic representation.*

PROOF. We first prove existence then uniqueness.

Step 1: We first prove existence using the technique of successive divisions we talked about before. Using the division algorithm, we form the remainders produced after successive divisions of a and its quotients by b:

$$a = q_0 \cdot b + a_0 \quad (\text{remainder } a_0), \quad q_0 = q_1 \cdot b + a_1, \quad (\text{remainder } a_1),$$
$$q_1 = q_2 \cdot b + a_2, \quad (\text{remainder } a_2), \ldots, \quad q_{j-1} = q_j \cdot b + a_j, \quad (\text{remainder } a_j), \ldots$$

and so forth. By the division algorithm, we have $q_j \geq 0$, $0 \leq a_j < b$, and moreover, since $b > 1$ (that is, $b \geq 2$), from the equation $q_{j-1} = q_j \cdot b + a_j$ it is evident that as long as the quotient q_0 is positive, we have

$$a = q_0 \cdot b + a_0 \geq q_0 \cdot b \geq 2q_0 > q_0$$

and in general, as long as the quotient q_j is positive, we have

$$q_{j-1} = q_j \cdot b + a_j \geq q_j \cdot b \geq 2q_j > q_j.$$

These inequalities imply that $a > q_0 > q_1 > q_2 > \cdots \geq 0$ where the strict inequality $>$ holds as long as the quotients remain positive. Since there are only a numbers from 0 to $a - 1$, at some point the quotients must eventually reach zero. Let us say that $q_n = 0$ is the first time the quotients hit zero. If $n = 0$, then we have

$$a = 0 \cdot b + a_0,$$

so a has the b-adic representation a_0. Suppose that $n > 0$. Then we have

$$\begin{aligned} a = q_0 \cdot b + a_0 \quad (\text{remainder } a_0), \quad q_0 = q_1 \cdot b + a_1, \quad (\text{remainder } a_1),\\ q_1 = q_2 \cdot b + a_2, \quad (\text{remainder } a_2), \ldots, \; q_{n-1} = 0 \cdot b + a_n, \quad (\text{remainder } a_n), \end{aligned} \tag{2.17}$$

and we stop successive divisions once we get a_n. Combining the first and second equations in (2.17), we get

$$a = q_0 \cdot b + a_0 = (q_1 \cdot b + a_1)b + a_0 = q_1 \cdot b^2 + a_1 b + a_0.$$

Combining this equation with the third equation in (2.17) we get

$$a = (q_2 \cdot b + a_2) \cdot b^2 + a_1 b + a_0 = q_2 \cdot b^3 + a_2 \cdot b^2 + a_1 b + a_0.$$

Continuing this process (slang for "by use of induction") we eventually arrive at

$$\begin{aligned} a &= (0 \cdot b + a_n) \cdot b^n + a_{n-1} \cdot b^{n-1} + \cdots + a_1 \cdot b + a_0 \\ &= a_n \cdot b^n + a_{n-1} \cdot b^{n-1} + \cdots + a_1 \cdot b + a_0. \end{aligned}$$

This shows the existence of a b-adic representation of a.

Step 2: We now show that this representation is unique. Suppose that a has another such representation:

$$a = \sum_{k=0}^{n} a_k\, b^k = \sum_{k=0}^{m} c_k\, b^k, \tag{2.18}$$

where $0 \le c_k < b$ and $c_m \neq 0$. We first prove that $n = m$. Indeed, let's suppose that $n \neq m$, say $n < m$. Then,

$$a = \sum_{k=0}^{n} a_k\, b^k \le \sum_{k=0}^{n} (b-1)\, b^k = b^{n+1} - 1.$$

Since $n < m \Longrightarrow n + 1 \le m$, we have $b^{n+1} \le b^m$, so

$$a \le b^m - 1 < b^m \le c_m \cdot b^m \le \sum_{k=0}^{m} c_k\, b^k = a \quad \Longrightarrow \quad a < a.$$

This contradiction shows that $n = m$. Now let us assume that some digits in the expressions for a differ; let p be the largest integer such that a_p differs from the corresponding c_p, say $a_p < c_p$. Since $a_p < c_p$ we have $a_p - c_p \le -1$. Now subtracting the two expressions for a in (2.18), we obtain

$$\begin{aligned} 0 = a - a = \sum_{k=0}^{p} (a_k - c_k) b^k &= \sum_{k=0}^{p-1} (a_k - c_k) b^k + (a_p - c_p) b^p \\ &\le \sum_{k=0}^{p-1} (b-1) b^k + (-b^p) = (b^p - 1) - b^p = -1, \end{aligned}$$

a contradiction. Thus, the two representation of a must be equal. □

Lastly, we remark that if a is negative, then $-a$ is positive, so $-a$ has a b-adic representation. The negative of this representation is by definition the b-adic representation of a.

Exercises 2.5.

1. In this exercise we consider the base twelve or **duodecimal system**. For this system we need two more digit symbols for eleven and twelve. Let α denote ten and β denote eleven. Then the digits for the duodecimal system are $0, 1, 2, \ldots, 9, \alpha, \beta$.
   (a) In the duodecimal system, what is the symbol for twelve, twenty-two, twenty-three, one hundred thirty-one?
   (b) What numbers do the following symbols represent? $\alpha\alpha\alpha$, 12, and $2\beta\beta 1$.

2. In the following exercises, we shall establish the validity of grade school divisibility "tricks", cf. [**112**]. Let $a = a_n a_{n-1} \dots a_0$ be the decimal (= base 10) representation of a natural number a. Let us first consider divisibility by $2, 5, 10$.
   (a) Prove that a is divisible by 10 if and only if $a_0 = 0$.
   (b) Prove that a is divisible by 2 if and only if a_0 is even.
   (c) Prove that a is divisible by 5 if and only if $a_0 = 0$ or $a_0 = 5$.
3. We now consider 4 and 8.
   (a) Prove that a is divisible by 4 if and only if the number $a_1 a_0$ (written in decimal notation) is divisible by 4.
   (b) Prove that a is divisible by 8 if and only if $a_2 a_1 a_0$ is divisible by 8.
4. We consider divisibility by $3, 6, 9$. (Unfortunately, there is no slick test for divisibility by 7.) Suggestion: Before considering these tests, prove that $10^k - 1$ is divisible by 9 for any nonnegative integer k.
   (a) Prove that a is divisible by 3 if and only if the sum of the digits (that is, $a_n + \cdots + a_1 + a_0$) is divisible by 3.
   (b) Prove that a is divisible by 6 if and only if a is even and the sum of the digits is divisible by 3.
   (c) Prove that a is divisible by 9 if and only if the sum of the digits is divisible by 9.
5. Prove that a is divisible by 11 if and only if the difference between the sums of the even and odd digits:

$$(a_0 + a_2 + a_4 + \cdots) - (a_1 + a_3 + a_5 + \cdots) = \sum_{k=0}^{n} (-1)^k a_k$$

   is divisible by 11. Suggestion: First prove that $10^{2k} - 1$ and $10^{2k+1} + 1$ are each divisible by 11 for any nonnegative integer k.
6. Using the idea of modular arithmetic from Problem 6 in Exercises 2.4, one can easily deduce the above "tricks". Take for example the "9 trick" and "11 trick".
   (a) Show that $10^k \equiv 1 \pmod 9$ for any $k = 0, 1, 2, \ldots$. Using this fact, prove that a is divisible by 9 if and only if the sum of the digits of a is divisible by 9.
   (b) Show that $10^k \equiv (-1)^k \pmod{11}$ for any $k = 0, 1, 2, \ldots$. Using this fact, prove that a is divisible by 11 if and only if the difference between the sums of the even and odd digits of a is divisible by 11.

2.6. Real numbers: Rational and "mostly" irrational

Imagine a world where you couldn't share half a cookie with your friend or where you couldn't buy a quarter pound of cheese at the grocery store; this is a world without rational numbers. In this section we discuss rational and real numbers and we shall discover, as the Greeks did 2500 years ago, that the rational numbers are not sufficient for the purposes of geometry. In particular, a world with only rational numbers is a world in which you couldn't measure the circumference of a circular swimming pool. Irrational numbers make up the missing rational lengths. We shall discover in the next few sections that there are vastly, immensely, incalculably (any other synonyms I missed?) more irrational numbers than rational numbers.

2.6.1. The real and rational numbers. The set of real numbers is denoted by $\mathbb{R}$. The reader is certainly familiar with the following arithmetic properties of real numbers (in what follows, a, b, c denote real numbers):

Addition satisfies

(A1) $a + b = b + a$; (commutative law)
(A2) $(a + b) + c = a + (b + c)$; (associative law)

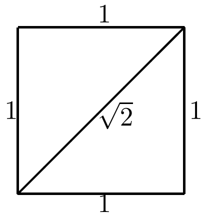


FIGURE 2.4. The Greek's discovery of irrational numbers.

(A3) there is a real number denoted by 0 "zero" such that

$$a + 0 = a = 0 + a; \quad \text{(existence of additive identity)}$$

(A4) for each a there is a real number denoted by the symbol $-a$ such that

$$a + (-a) = 0 \quad \text{and} \quad (-a) + a = 0. \quad \text{(existence of additive inverse)}$$

Multiplication satisfies

(M1) $a \cdot b = b \cdot a$; (commutative law)
(M2) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$; (associative law)
(M3) there is a real number denoted by 1 "one" such that

$$1 \cdot a = a = a \cdot 1; \quad \text{(existence of multiplicative identity)}$$

(M4) for $a \neq 0$ there is a real number denoted by the symbol a^{-1} such that

$$a \cdot a^{-1} = 1 \quad \text{and} \quad a^{-1} \cdot a = 1. \quad \text{(existence of multiplicative inverse)}$$

As with the integers, the $\cdot$ is sometimes dropped and the associative laws imply that expressions such as $a + b + c$ or abc make sense without using parentheses. Addition and multiplication are related by

(D) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$. (distributive law)

Of these arithmetic properties, the only additional property listed that was not listed for integers is (**M4**), the existence of a multiplicative inverse for each nonzero real number. As usual, we denote

$$a + (-b) = (-b) + a \quad \text{by} \quad b - a$$

and

$$a \cdot b^{-1} = b^{-1} \cdot a \quad \text{by} \quad a/b \quad \text{or} \quad \frac{a}{b}.$$

The positive real numbers, denoted $\mathbb{R}^+$, are closed under addition, multiplication, and has the following property: Given any real number a, exactly one of the following "positivity" properties hold:

(P) a is a positive real number, $a = 0$, or $-a$ is a positive real number.

A set together with operations of addition and multiplication that satisfy properties (**A1**) – (**A4**), (**M1**) – (**M4**), and (**D**) is called a **field**; essentially a field is just a set of objects closed under addition, multiplication, subtraction, and division (by nonzero elements). If in addition, the set has a "positive set" closed under addition and multiplication satisfying (**P**), then the set is called an **ordered field**.

A **rational number** is a number that can be written in the form a/b where a and b are integers with $b \neq 0$ and the set of all such numbers is denoted by $\mathbb{Q}$.

We leave the reader to check that the rational numbers also form an ordered field. Thus, both the real numbers and the rational numbers are ordered fields. Now what is the difference between the real and rational numbers? The difference was discovered more than 2500 years ago by the Greeks, who found out that the length of the diagonal of a unit square, which according to the Pythagorean theorem is $\sqrt{2}$, is not a rational number (see Theorem 2.23). Because this length is not a rational number, the Greeks called a number such as $\sqrt{2}$ **irrational**.[9] Thus, there are "gaps" in the rational numbers. Now it turns out that *every* length is a real number. This fact is known as the completeness axiom of the real numbers. Thus, the real numbers have no "gaps". To finish up the list of axioms for the real numbers, we state this completeness axiom now but we leave the terms in the axiom undefined until Section 2.7 (so don't worry if some of these words seem foreign).

(C) (**Completeness axiom of the real numbers**) Every nonempty set of real numbers that is bounded above has a supremum, that is, a least upper bound.

We shall *assume* the existence of a set $\mathbb{R}$ such that $\mathbb{N} \subseteq \mathbb{R}^+$, $\mathbb{Z} \subseteq \mathbb{R}$, and $\mathbb{R}$ satisfies all the arithmetic, positivity, and completeness properties listed above.[10] All theorems that we prove in this textbook are based on this assumption.

2.6.2. Proofs of well-known high school rules. Since the real numbers satisfy the same arithmetic properties as the natural and integer numbers, the same proofs as in Section 2.1 and 2.3 prove the uniqueness of additive identities and inverses, rules of sign, properties of zero and one (in particular, the uniqueness of the multiplicative identity), etc ....

Also, the real numbers are ordered in the same way as the integers. Given any real numbers a and b exactly one of the following holds:

(O1) $a = b$;
(O2) $a < b$, which means that $b - a$ is a positive real number;
(O3) $b < a$, which means that $-(b - a)$ is a positive real number.

Just as for integers, we can define $\leq$, $>$, and $\geq$ and (**O3**) is just that $a - b$ is a positive real number. One can define the absolute value of a real number in the exact same way as it is defined for integers. Since the real numbers satisfy the same order properties as the integers, the same proofs as in Section 2.3 prove the inequality rules, absolute value rules, etc ..., for real numbers. Using the inequality rules we can prove the following well-known fact from high school: if $a > 0$, then $a^{-1} > 0$. Indeed, by definition of a^{-1}, we have $a \cdot a^{-1} = 1$. Since $1 > 0$ (recall that $1 \in \mathbb{N} \subseteq \mathbb{R}^+$) and $a > 0$, we have positive $\times\, a^{-1} =$ positive; the only way this is possible is if $a^{-1} > 0$ by the inequality rules. Here are other high school facts that can be proved using the inequality rules: If $0 < a < 1$, then $a^{-1} > 1$ and if $a > 1$, then $a^{-1} < 1$. Indeed, if $a < 1$ with a positive, then multiplying by $a^{-1} > 0$, we obtain

$$a \cdot (a^{-1}) < 1 \cdot (a^{-1}) \quad \Longrightarrow \quad 1 < a^{-1}.$$

[9] *The idea of the continuum seems simple to us. We have somehow lost sight of the difficulties it implies ... We are told such a number as the square root of 2 worried Pythagoras and his school almost to exhaustion. Being used to such queer numbers from early childhood, we must be careful not to form a low idea of the mathematical intuition of these ancient sages; their worry was highly credible. Erwin Schrödinger (1887–1961).*

[10] For simplicity we assumed that $\mathbb{N} \subseteq \mathbb{R}^+$ (in particular, all natural numbers are positive by assumption) and $\mathbb{Z} \subseteq \mathbb{R}$, but it is possible to define $\mathbb{N}$ and $\mathbb{Z}$ within $\mathbb{R}$. Actually, one only needs to define $\mathbb{N}$ for then we can put $\mathbb{Z} := \mathbb{N} \cup \{0\} \cup (-\mathbb{N})$.

Similarly, if $1 < a$, then multiplying through by $a^{-1} > 0$, we get $a^{-1} < 1$.

Here are some more high school facts.

THEOREM 2.20 (**Uniqueness of multiplicative inverse**). *If a and b are real numbers with $a \neq 0$, then $x \cdot a = b$ if and only if $x = ba^{-1} = b/a$. In particular, setting $b = 1$, the only x that satisfies the equation $x \cdot a = 1$ is $x = a^{-1}$. Thus, each real number has only one multiplicative inverse.*

PROOF. If $x = b \cdot a^{-1}$, then

$$(ba^{-1}) \cdot a = b(a^{-1}a) = b \cdot 1 = b,$$

so the real number $x = b/a$ solves the equation $x \cdot a = b$. Conversely, if x satisfies $x \cdot a = b$, then

$$ba^{-1} = (x \cdot a)a^{-1} = x \cdot (a\,a^{-1}) = x \cdot 1 = x.$$

□

Recall that $x \cdot 0 = 0$ for any real number x. (This is Theorem 2.12 in the real number case.) In particular, 0 has no multiplicative inverse (there is no real number "0^{-1}" such that $0 \cdot 0^{-1} = 1$); thus, the high school saying: "You can't divide by zero."

THEOREM 2.21 (**Fraction rules**). *For $a, b, c, d \in \mathbb{R}$, the following fraction rules hold (all denominators are assumed to be nonzero):*

$$(1)\ \ \frac{a}{a} = 1,\ \frac{a}{1} = a, \qquad (2)\ \ \frac{a}{-b} = -\frac{a}{b}$$

$$(3)\ \ \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad (4)\ \ \frac{a}{b} = \frac{ac}{bc},$$

$$(5)\ \ \frac{1}{a/b} = \frac{b}{a}, \qquad (6)\ \ \frac{a/b}{c/d} = \frac{a}{b} \cdot \frac{d}{c} = \frac{ad}{bc}, \qquad (7)\ \ \frac{a}{b} \pm \frac{c}{d} = \frac{ad \pm bc}{bd}.$$

PROOF. The proofs of these rules are really very elementary, so we only prove *(1)–(3)* and leave *(4)–(7)* to you in Problem 1.

We have $a/a = a \cdot a^{-1} = 1$ and since $1 \cdot 1 = 1$, by uniqueness of the multiplicative inverses, we have $1^{-1} = 1$ and therefore $a/1 = a \cdot 1^{-1} = a \cdot 1 = a$.

To prove *(2)*, note that by our rules of sign,

$$(-b) \cdot (-b^{-1}) = b \cdot b^{-1} = 1$$

and therefore by uniqueness of multiplicative inverses, we must have $(-b)^{-1} = -b^{-1}$. Thus, $a/(-b) := a \cdot (-b)^{-1} = a \cdot -b^{-1} = -a \cdot b^{-1} = -a/b$.

To prove *(3)*, observe that $b \cdot d \cdot b^{-1} \cdot d^{-1} = (bb^{-1}) \cdot (dd^{-1}) = 1 \cdot 1 = 1$, so by uniqueness of multiplicative inverses, $(bd)^{-1} = b^{-1}d^{-1}$. Thus,

$$\frac{a}{b} \cdot \frac{c}{d} = a \cdot b^{-1} \cdot c \cdot d^{-1} = a \cdot c \cdot b^{-1} \cdot d^{-1} = ac \cdot (bd)^{-1} = \frac{ac}{bd}.$$

□

We already know that what a^n means for $n = 0, 1, 2, \ldots$. For negative integers, we define powers by

$$\boxed{a^{-n} := \frac{1}{a^n}, \quad a \neq 0,\ n = 1, 2, 3, \ldots.}$$

Here are the familiar power rules.

THEOREM 2.22 (**Power rules**). *For $a, b \in \mathbb{R}$ and for integers m, n,*

$$a^m \cdot a^n = a^{m+n}; \quad a^m \cdot b^m = (ab)^m; \quad (a^m)^n = a^{mn},$$

provided that the individual powers are defined (e.g. a and b are nonzero if an exponent is negative). If n is a natural number and $a, b \geq 0$, then

$$a < b \quad \text{if and only if} \quad a^n < b^n.$$

In particular, $a \neq b$ (both a, b nonnegative) if and only if $a^n \neq b^n$.

PROOF. We leave the proof of the first three rules to the reader since we already dealt with proving such rules in the problems of Section 2.2. Consider the last rule. Let $n \in \mathbb{N}$ and let $a, b \geq 0$ be not both zero (if both are zero, then $a = b$ and $a^n = b^n$ and there is nothing to prove). Observe that

$$(b - a) \cdot c = b^n - a^n, \quad \text{where } c = b^{n-1} + b^{n-2}\, a + \cdots + b\, a^{n-2} + a^{n-1}, \tag{2.19}$$

which is verified by multiplying out:

$$\begin{aligned}(b - a)\,(b^{n-1} + b^{n-2}\, a + \cdots + a^{n-1}) &= (b^n + b^{n-2}\, a^2 + b^{n-1}\, a + \cdots + ba^{n-1}) \\ &\quad - (b^{n-1}a + b^{n-2}\, a^2 + \cdots + b\, a^{n-1} + a^n) = b^n - a^n.\end{aligned}$$

The formula for c (and the fact that $a, b \geq 0$ are not both zero) shows that $c > 0$. Therefore, the equation $(b - a) \cdot c = b^n - a^n$ shows that $(b - a) > 0$ if and only if $b^n - a^n > 0$. Therefore, $a < b$ if and only if $a^n < b^n$. □

If n is a natural number, then the n-th **root** of a real number a is a real number b such that $b^n = a$, if such a number b exists. For $n = 2$, we usually call b a **square root** and if $n = 3$, a **cube root**.

Example 2.18. -3 is a square root of 9 since $(-3)^2 = 9$. Also, 3 is a square root of 9 since $3^2 = 9$. Here's a puzzle: Which two real numbers are their own nonnegative square roots?

If $a \geq 0$, then according to last power rule in Theorem 2.22, a can have at most one nonnegative n-th root. In Section 2.7 we shall prove that *any* nonnegative real number has a unique nonnegative n-th root. We denote this unique n-th root by $\sqrt[n]{a}$ or $a^{1/n}$. If $n = 2$, we always write $\sqrt{a}$ or $a^{1/2}$ instead of $\sqrt[2]{a}$.

We now show that "most" real numbers are not rational numbers, that is, ratios of integers. These examples will convince the reader that there are many "gaps" in the rational numbers and the importance of irrational numbers to real life. For the rest of this section, we shall *assume* that $\sqrt[n]{a}$ exists for any $a \geq 0$ (to be proved in Section 2.7) and we shall assume basic facts concerning the trig and log functions (to be proved in Sections 4.7 and 4.6, respectively) . We make these assumptions only to present interesting examples that will convince you without a shadow of a doubt that irrational numbers are indispensable in mathematics.

2.6.3. Irrational roots and the rational zeros theorem. We begin by showing that $\sqrt{2}$ is not rational. Before proving this, we establish some terminology. We say that a rational number a/b is in **lowest terms** if a and b do not have common prime factors in their prime factorizations. By Property *(4)* of the fraction rules, we can always "cancel" common factors to put a rational number in lowest terms.

THEOREM 2.23 (**Irrationality of $\sqrt{2}$**). *$\sqrt{2}$ is not rational.*

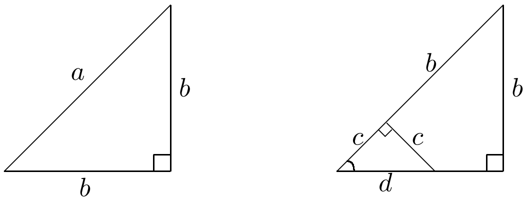


FIGURE 2.5. On the right, we measure the length b along the longer side a and we draw a perpendicular from side a to the shorter side b. We get a new isosceles triangle with sides d, c, c. (The smaller triangle is similar to the original one because it has a 90° angle just like the original one does and it shares an angle, the lower left corner, with the original one.)

PROOF. We provide three proofs, the first one is essentially a version (the version?) of the original geometric Pythagorean proof while the second one is a real analysis version of the same proof! The third proof is the "standard" proof in this business. (See also Problems 6 and 7.)

Proof I: (Cf. [**8**] for another version.) This proof is not to be considered rigorous! We only put this proof here for historical purposes and because we shall make this proof completely rigorous in **Proof II** below. We assume common facts from high school geometry, in particular, similar triangles.

Suppose, by way of contradiction, that $\sqrt{2} = a/b$ where $a, b \in \mathbb{N}$. Then $a^2 = 2b^2 = b^2 + b^2$, so by the Pythagorean theorem, the isosceles triangle with sides a, b, b is a right triangle (see Figure 2.5). Hence, there is an isosceles right triangle whose lengths are (of course, positive) integers. By taking a smaller triangle if necessary, we may assume that a, b, b are the lengths of the smallest such triangle. We shall derive a contradiction by producing another isosceles right triangle with integer lengths and a smaller hypotonus. In fact, consider the triangle d, c, c drawn in Figure 2.5. Note that $a = b + c$ so $c = a - b \in \mathbb{Z}$. To see that $d \in \mathbb{Z}$, observe that since the ratio of corresponding sides of similar triangles are in proportion, we have

$$\frac{d}{c} = \frac{a}{b} \quad \Longrightarrow \quad d = \frac{a}{b} \cdot c = \frac{a}{b}(a - b) = \frac{a^2}{b} - a = 2b - a, \tag{2.20}$$

where we used that $a^2 = b^2 + b^2 = 2b^2$. Therefore, $d = 2b - a \in \mathbb{Z}$ as well. Thus, we have indeed produced a smaller isosceles right triangle with integer lengths.

Proof II: (Cf. [**218**, p. 39], [**143**], [**194**].) We now make **Proof I** rigorous. Suppose that $\sqrt{2} = a/b$ $(a, b \in \mathbb{N})$. By well-ordering, we may assume that a is the smallest positive numerator that $\sqrt{2}$ can have as a fraction; explicitly,

$$a = \text{least element of } \left\{ n \in \mathbb{N} \,;\, \sqrt{2} = \frac{n}{m} \text{ for some } m \in \mathbb{Z} \right\}.$$

Motivated by (2.20), we *claim* that

$$\sqrt{2} = \frac{d}{c} \quad \text{where} \quad d = 2b - a, c = a - b \text{ are integers with } d \in \mathbb{N} \text{ and } d < a. \tag{2.21}$$

Once we prove this claim, we contradict the minimality of a. Of course, the facts in (2.21) were derived from Figure 2.5 geometrically, but now we actually prove these facts! First, to prove that $\sqrt{2} = d/c$, we simply compute:

$$\begin{aligned}\frac{d}{c} = \frac{2b-a}{a-b} = \frac{2-a/b}{a/b-1} = \frac{2-\sqrt{2}}{\sqrt{2}-1} &= \frac{2-\sqrt{2}}{\sqrt{2}-1} \cdot \frac{\sqrt{2}+1}{\sqrt{2}+1} \\ &= \frac{2\sqrt{2}+2-(\sqrt{2})^2-\sqrt{2}}{(\sqrt{2})^2-1} = \frac{\sqrt{2}}{1} = \sqrt{2}.\end{aligned}$$

To prove that $0 < d < a$, note that since $1 < 2 < 4$, that is, $1^2 < (\sqrt{2})^2 < 2^2$, by the (last statement of the) power rules in Theorem 2.22, we have $1 < \sqrt{2} < 2$, or $1 < a/b < 2$. Multiplying by b, we get $b < a < 2b$, which implies that

$$d = 2b - a > 0 \quad \text{and} \quad d = 2b - a < 2a - a = a. \tag{2.22}$$

Therefore, $d \in \mathbb{N}$ and $d < a$ and we get our a contradiction.

Proof III: The following proof is the classic proof. We first establish the fact that the square of an integer has the factor 2 if and only if the integer itself has the factor 2. A quick way to prove this fact is using the fundamental theorem of arithmetic: The factors of m^2 are exactly the squares of the factors of m. Therefore, m^2 has a prime factor p if and only if m itself has the prime factor p. In particular, m^2 has the prime factor 2 if and only if m has the factor 2, which establishes our fact. A proof without using the fundamental theorem goes as follows. An integer is either even or odd, that is, is of the form $2n$ or $2n+1$ where n is the quotient of the integer when divided by 2. The equations

$$(2n)^2 = 4n^2 = 2(2n^2)$$
$$(2n+1)^2 = 4n^2 + 4n + 1 = 2(2n^2 + 2n) + 1$$

confirm the asserted fact. Now suppose that $\sqrt{2}$ were a rational number, say

$$\sqrt{2} = \frac{a}{b},$$

where a/b is in lowest terms. Squaring this equation we get

$$2 = \frac{a^2}{b^2} \quad \Longrightarrow \quad a^2 = 2b^2.$$

The number $2b^2 = a^2$ has the factor 2, so a must have the factor 2. Therefore, $a = 2c$ for some integer c. Thus,

$$(2c)^2 = 2b^2 \quad \Longrightarrow \quad 4c^2 = 2b^2 \quad \Longrightarrow \quad 2c^2 = b^2.$$

The number $2c^2 = b^2$ has the factor 2, so b must also have the factor 2. Thus, we have showed that a and b both have the factor 2. This contradicts the assumption that a and b have no common factors. □

The following theorem gives another method to prove the irrationality of $\sqrt{2}$ and also many other numbers. Recall that a (real-valued) **n-th degree polynomial** is a function $p(x) = a_n x^n + \cdots + a_1 + x + a_0$, where $a_k \in \mathbb{R}$ for each k and with the **leading coefficient** $a_n \neq 0$.

Theorem 2.24 (**Rational zeros theorem**). *If a polynomial equation with integral coefficients,*

$$c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0, \quad c_n \neq 0,$$

where the c_k's are integers, has a nonzero rational solution a/b where a/b is in lowest terms, then a divides c_0 and b divides c_n.

Proof. Suppose that a/b is a rational solution of our equation with a/b in lowest terms. Being a solution, we have

$$c_n\left(\frac{a}{b}\right)^n + c_{n-1}\left(\frac{a}{b}\right)^{n-1} + \cdots + c_1\left(\frac{a}{b}\right) + c_0 = 0.$$

Multiplying both sides by b^n, we obtain

$$c_na^n + c_{n-1}a^{n-1}\,b + \cdots + c_1a\,b^{n-1} + c_0b^n = 0. \tag{2.23}$$

Bringing everything to the right except for c_na^n and factoring out a b, we find

$$\begin{aligned} c_na^n &= -c_{n-1}a^{n-1}\,b - \cdots - c_1a\,b^{n-1} - c_0b^n \\ &= b(-c_{n-1}a^{n-1} - \cdots - c_1a\,b^{n-2} - c_0b^{n-1}). \end{aligned}$$

This formula shows that every prime factor of b occurs in the product c_na^n. By assumption, a and b have no common prime factors and hence every prime factor of b must occur in c_n. This shows that b divides c_n.

We now rewrite (2.23) as

$$\begin{aligned} c_0b^n &= -c_na^n - c_{n-1}a^{n-1}\,b - \cdots - c_1a\,b^{n-1} \\ &= a(-c_na^{n-1} - c_{n-1}a^{n-2} - \cdots - c_1b^{n-1}). \end{aligned}$$

This formula shows that every prime factor of a occurs in the product c_0b^n. However, since a and b have no common prime factors, we conclude that every prime factor of a occurs in c_0, which implies that a divides c_0. This completes our proof. □

Example 2.19. (**Irrationality of $\sqrt{2}$, Proof IV**) Observe that $\sqrt{2}$ is a solution of the polynomial equation $x^2 - 2 = 0$. The rational zeros theorem implies that if the equation $x^2 - 2 = 0$ has a rational solution, say a/b in lowest terms, then a must divide $c_0 = -2$ and b must divide $c_2 = 1$. It follows that a can equal ± 1 or ± 2 and b can only be ± 1. Therefore, the only rational solutions of $x^2 - 2 = 0$, if any, are $x = \pm 1$ or $x = \pm 2$. However,

$$(\pm 1)^2 - 2 = -1 \neq 0 \quad \text{and} \quad (\pm 2)^2 - 2 = 2 \neq 0,$$

so $x^2 - 2 = 0$ has no rational solutions. Therefore $\sqrt{2}$ is not rational.

A similar argument using the equation $x^n - a = 0$ proves the the following corollary.

Corollary 2.25. *The n-th root $\sqrt[n]{a}$, where a and n are positive integers, is either irrational or an integer; if it is an integer, then a is the n-th power of an integer.*

2.6.4. Irrationality of trigonometric numbers. Let $0 < \theta < 90°$ be an angle whose measurement in degrees is rational. Following [**142**], we shall prove that $\cos\theta$ is irrational except when $\theta = 60°$, in which case

$$\cos 60° = \frac{1}{2}.$$

The proof of this result is based on the rational zero theorem and Lemma 2.26 below. See Problem 5 for corresponding statements for sine and tangent. Of course, at this point, and only for purposes of illustration, we have to assume basic knowledge of the trigonometric functions. In Section 4.7 we shall define these function rigourously and establish their usual properties.

LEMMA 2.26. *For any natural number n, we can write $2\cos n\theta$ as an n-th degree polynomial in $2\cos\theta$ with integer coefficients and with leading coefficient one.*

PROOF. We need to prove that

$$2\cos n\theta = (2\cos\theta)^n + a_{n-1}(2\cos\theta)^{n-1} + \cdots + a_1(2\cos\theta) + a_0, \tag{2.24}$$

where the coefficients $a_{n-1}, a_{n-2}, \ldots, a_0$ are integers. For $n = 1$, we can write $2\cos\theta = (2\cos\theta)^1 + 0$, so our proposition holds for $n = 1$. To prove our result in general, we use the strong form of induction. Assume that our proposition holds for $1, 2, \ldots, n$. Before proceeding to show that our lemma holds for $n + 1$, we shall prove the identity

$$2\cos(n+1)\theta = \Big(2\cos n\theta\Big)\Big(2\cos\theta\Big) - 2\cos(n-1)\theta. \tag{2.25}$$

To verify this identity, consider the identities

$$\begin{aligned}\cos(\alpha+\beta) &= \cos\alpha\,\cos\beta - \sin\alpha\,\sin\beta\\ \cos(\alpha-\beta) &= \cos\alpha\,\cos\beta + \sin\alpha\,\sin\beta.\end{aligned}$$

Adding these equations, we obtain $\cos(\alpha+\beta) + \cos(\alpha-\beta) = 2\cos\alpha\,\cos\beta$, or

$$\cos(\alpha+\beta) = 2\cos\alpha\,\cos\beta - \cos(\alpha-\beta).$$

Setting $\alpha = n\theta$ and $\beta = \theta$, and then multiplying the result by 2, we get (2.25).

Now, since our lemma holds for $1, \ldots, n$, in particular, $2\cos(n-1)\theta$ can be written as an $(n-1)$-degree polynomial in $2\cos\theta$ with integer coefficients and with leading coefficient one and $2\cos n\theta$ can be written as an n-degree polynomial in $2\cos\theta$ with integer coefficients and with leading coefficient one. Substituting these polynomials into the right-hand side of the identity (2.25) shows that $2\cos(n+1)\theta$ can be expressed as an $(n+1)$-degree polynomial in $2\cos\theta$ with integer coefficients and with leading coefficient one. This proves our lemma. □

We are now ready to prove our main result.

THEOREM 2.27. *Let $0 < \theta < 90°$ be an angle whose measurement in degrees is rational. Then $\cos\theta$ is rational if and only if $\theta = 60°$.*

PROOF. If $\theta = 60°$, then we know that $\cos\theta = 1/2$, which is rational.

Assume now that θ is rational, say $\theta = a/b$ where a and b are natural numbers. Then choosing $n = b \cdot 360°$, we have $n\theta = b \cdot 360° \cdot (a/b) = a \cdot 360°$. Thus, $n\theta$ is a multiple of $360°$, so $\cos n\theta = 1$. Substituting $n\theta$ into the equation (2.24), we obtain

$$(2\cos\theta)^n + a_{n-1}(2\cos\theta)^{n-1} + \cdots + a_1(2\cos\theta) + a_0 - 2 = 0,$$

where the coefficients are integers. Hence, $2\cos\theta$ is a solution of the equation

$$x^n + a_{n-1}\,x^{n-1} + \cdots + a_1\,x + a_0 - 2 = 0.$$

By the rational zeros theorem, any rational solution of this equation must be an integer dividing -2. So, if $2\cos\theta$ is rational, then it must be an integer. Since $0 < \theta < 90°$ and cosine is strictly between 0 and 1 for these θ's, the only integer that $2\cos\theta$ can be is 1. Thus, $2\cos\theta = 1$ or $\cos\theta = 1/2$, and so θ must be $60°$. □

2.6.5. Irrationality of logarithmic numbers. Recall that the **(common) logarithm** to the base 10 of a real number a is defined to be the unique number x such that

$$10^x = a.$$

In Section 4.6 we define logarithms rigourously but for now, and only now, in order to demonstrate another interesting example of irrational numbers, we shall assume familiarity with such logarithms from high school. We also assume basic facts concerning powers that we'll prove in the next section.

THEOREM 2.28. *Let $r > 0$ be any rational number. Then $\log_{10} r$ is rational if and only if $r = 10^n$ where n is an integer, in which case*

$$\log_{10} r = n.$$

PROOF. If $r = 10^n$ where $n \in \mathbb{Z}$, then $\log_{10} r = n$, so $\log_{10} r$ is rational. Assume now that $\log_{10} r$ is rational; we'll show that $r = 10^n$ for some $n \in \mathbb{Z}$. We may assume that $r > 1$ because if $r = 1$, then $r = 10^0$, and we're done, and if $r < 1$, then $r^{-1} > 1$ and $\log_{10} r^{-1} = -\log_{10} r$ is rational, so we can get the $r < 1$ result from the $r > 1$ result. We henceforth assume that $r > 1$. Let $r = a/b$ where a and b are natural numbers with no common factors. Assume that $\log_{10} r = c/d$ where c and d are natural numbers with no common factors. Then $r = 10^{c/d}$, which implies that $r^d = 10^c$, or after setting $r = a/b$, we get $(a/b)^d = 10^c \Longrightarrow a^d = 10^c \cdot b^d$, or

$$a^d = 2^c \cdot 5^c \cdot b^d. \tag{2.26}$$

By assumption, a and b do not have any common prime factors. Hence, expressing a and b in the their prime factorizations in (2.26) and using the fundamental theorem of arithmetic, we see that the only way (2.26) can hold is if b has no prime factors (that is, $b = 1$) and a can only have the prime factors 2 and 5. Thus,

$$a = 2^m \cdot 5^n \quad \text{and} \quad b = 1$$

for some nonnegative integers m and n. Now according to (2.26),

$$2^{md} \cdot 5^{nd} = 2^c \cdot 5^c.$$

Again by the fundamental theorem of arithmetic, we must have $md = c$ and $nd = c$. Now c and d have no common factors, so $d = 1$, and therefore $m = c = n$. This, and the fact that $b = 1$, proves that

$$r = \frac{a}{b} = \frac{a}{1} = 2^m \cdot 5^n = 2^n \cdot 5^n = 10^n.$$

□

In the following exercises, assume that square roots and cube roots exist for nonnegative real numbers; again, this fact will be proved in the next section.

EXERCISES 2.6.

1. Prove properties *(4)–(7)* in the "Fraction rules" theorem.
2. Let a be any positive real number and let n, m be nonnegative integers with $m < n$. If $0 < a < 1$, prove that $a^n < a^m$ and if $a > 1$, prove that $a^m < a^n$.
3. Let α be any irrational number. Prove that $-\alpha$ and α^{-1} are irrational. If r is any nonzero rational number, prove that the addition, subtraction, multiplication, and division of α and r are again irrational. As an application of this result, deduce that

$$-\sqrt{2}, \qquad \frac{1}{\sqrt{2}}, \qquad \sqrt{2}+1, \qquad 4-\sqrt{2}, \qquad 3\sqrt{2}, \qquad \frac{\sqrt{2}}{10}, \qquad \frac{7}{\sqrt{2}}$$

are each irrational.

4. In this problem we prove that various numbers are irrational.
   (a) Prove that $\sqrt{6}$ is irrational using the **Proof III** in Theorem 2.23. From the fact that $\sqrt{6}$ is irrational, and without using any irrationality facts concerning $\sqrt{2}$ and $\sqrt{3}$, prove that $\sqrt{2}+\sqrt{3}$ is irrational. Suggestion: To prove that $\sqrt{2}+\sqrt{3}$ is irrational, consider its square.
   (b) Now prove that $\sqrt{2}+\sqrt{3}$ is irrational using the rationals zero theorem. Suggestion: Let $x=\sqrt{2}+\sqrt{3}$, then show that $x^4-10x^2+1=0$.
   (c) Using the rationals zero theorem, prove that $(2\sqrt[3]{6}+7)/3$ is irrational and $\sqrt[3]{2}-\sqrt{3}$ is irrational. (If $x=\sqrt[3]{2}-\sqrt{3}$, you should end up with a sixth degree polynomial equation for x for which you can apply the rationals zero theorem.)
5. In this problem we look at irrational values of sine and tangent. Let $0<\theta<90°$ be an angle whose measurement in degrees is rational. You may assume *any* knowledge of the trigonometric functions and their identities.
   (a) Prove that $\sin\theta$ is rational if and only if $\theta=30°$, in which case $\sin 30°=1/2$. Suggestion: Do *not* try to imitate the proof of Theorem 2.27, instead use a trig identity to write sine in terms of cosine.
   (b) Prove that $\tan\theta$ is rational if and only if $\theta=45°$, in which case $\tan\theta=1$. Suggestion: Use the identity $\cos 2\theta=\frac{1-\tan^2\theta}{1+\tan^2\theta}$.
6. (Cf. [**43**]) (**Irrationality of $\sqrt{2}$, Proof V**) This proof is similar to the algebraic Pythagorean proof of Theorem 2.23. Assume that $\sqrt{2}$ is rational.
   (i) Show that there is a smallest natural number n such that $n\sqrt{2}$ is an integer.
   (ii) Show that $m=n\sqrt{2}-n=n(\sqrt{2}-1)$ is a natural number smaller than n.
   (iii) Finally, show that $m\sqrt{2}$ is an integer, which contradicts the fact that n was the smallest natural number having this property.
7. (**Irrationality of $\sqrt{2}$, Proof VI**) Here's a proof due to Marcin Mazur [**148**].
   (i) Show that $\sqrt{2}=\frac{-4\sqrt{2}+6}{3\sqrt{2}-4}$.
   (ii) Now suppose that $\sqrt{2}=a/b$ $(a,b\in\mathbb{N})$ where a is the smallest positive numerator that $\sqrt{2}$ can have as a fraction. Using the formula in (i), derive a contradiction as in the algebraic Pythagorean proof of Theorem 2.23.

2.7. The completeness axiom of $\mathbb{R}$ and its consequences

The completeness axiom of the real numbers essentially states that the real numbers have no "gaps". As discovered in the previous section, this property is quite in contrast to the rational numbers that have many "gaps". In this section we discuss the completeness axiom and its consequences. Another consequence of the completeness axiom is that the real numbers is uncountable while the rationals are countable, but we leave this breath-taking subject for Section 2.10.

2.7.1. The completeness axiom.

Before discussing the completeness axiom, we need to talk about lower and upper bounds of sets.

A set $A\subseteq\mathbb{R}$ is said to be **bounded above** if there is a real number b larger than any number in A in the sense that for each a in A we have $a\le b$. Any such number b, if such exists, is called an **upper bound** for A. Suppose that b is an upper bound for A. Then b is called the **least upper bound** or **supremum** for A if b is just that, the least upper bound for A, in the sense that it is less than any other upper bound for A. This supremum, if it exists, is denoted by $\sup A$. We shall use both terminologies "least upper bound" and "supremum" interchangeably although we shall use least upper bound more often.

Example 2.20. Consider the interval $I = [0, 1)$. This interval is bounded above by, for instance, 1, 3/2, 22/7, 10, 1000, etc. In fact, any upper bound for I is just a real number greater than or equal to 1. The least upper bound is 1 since 1 is the smallest upper bound. Note that $1 \notin I$.

Example 2.21. Now let $J = (0, 1]$. This set is also bounded above, and any upper bound for J is as before, just a real number greater than or equal to 1. The least upper bound is 1. In this case, $1 \in J$.

These examples show that the supremum of a set, if it exists, may or may not belong to the set.

Example 2.22. $\mathbb{Z}$ is not bounded above (see Lemma 2.34) nor is the set $(0, \infty)$.

We summarize: Let $A \subseteq \mathbb{R}$ be bounded above. Then a number b is the least upper bound or supremum for A means two things concerning b:

(L1) for all a in A, $a \le b$ — this just means that b is an upper bound for A;

(L2) if c is an upper bound for A, then $b \le c$ — this just means that b is the least, or smallest, upper bound for A.

Instead of **(L2)** it is sometimes convenient to substitute the following.

(L2′) if $c < b$, then for some a in A we have $c < a$ — this just means that any number c smaller than b cannot be an upper bound for A, which is to say, there is no upper bound for A that is smaller than b.

(L2′) is just the contrapositive of **(L2)** — do you see why? We can also talk about lower bounds. A set $A \subseteq \mathbb{R}$ is said to be **bounded below** if there is a real number b smaller than any number in A in the sense that for each a in A we have $b \le a$. Any such number b, if such exists, is called a **lower bound** for A. If b is a lower bound for A, then b is called the **greatest lower bound** or **infimum** for A if b is just that, the greatest lower bound for A, in the sense that it is greater than any other lower bound for A. This infimum, if it exists, is denoted by $\inf A$. We shall use both terminologies "greatest lower bound" and "infimum" interchangeably although we shall use greatest lower bound more often.

Example 2.23. The sets $I = [0, 1)$ and $J = (0, 1]$ are both bounded below (by e.g. 0, $-1/2$, -1, -1000, etc.) and in both cases the greatest lower bound is 0.

Thus, the infimum of a set, if it exists, may or may not belong to the set.

Example 2.24. $\mathbb{Z}$ (see Lemma 2.34) and $(-\infty, 0)$ are not bounded below.

We summarize: Let $A \subseteq \mathbb{R}$ be bounded below. Then a number b is the greatest lower bound or infimum for A means two things concerning b:

(G1) for all a in A, $b \le a$ — this just means that b is a lower bound for A,

(G2) if c is a lower bound for A, then $c \le b$ — this just means that b is the greatest lower bound for A.

Instead of **(G2)** it is sometimes convenient to substitute its contrapositive.

(G2′) if $b < c$, then for some a in A we have $a < c$ — this just means that any number c greater than b cannot be a lower bound for A, which is to say, there is no lower bound for A that is greater than b.

In the examples given so far (e.g. the intervals I and J), we have shown that if a set has an upper bound, then it has a least upper bound. This is a general phenomenon, called the **completeness axiom** of the real numbers:

(C) (**Completeness axiom of the real numbers**) Every nonempty set of real numbers that is bounded above has a supremum, that is, a least upper bound.

As stated in the last section, we *assume* that $\mathbb{R}$ has this property. Using the following lemma, we can prove the corresponding statement for infimums.

LEMMA 2.29. *If A is nonempty and bounded below, then $-A := \{-a\,;\, a \in A\}$ is nonempty and bounded above, and* $\inf A = -\sup(-A)$ *in the sense that* $\inf A$ *exists and this formula for* $\inf A$ *holds.*

PROOF. Since A is nonempty and bounded below, there is a real number b such that $b \leq a$ for all a in A. Therefore, $-a \leq -b$ for all a in A, and hence the set $-A$ is bounded above by $-b$. By the completeness axiom, $-A$ has a least upper bound, which we denote by b. Our lemma is finished once we show that $-b$ is the greatest lower bound for A. To see this, we know that $-a \leq b$ for all a in A and so, $-b \leq a$ for all a in A. Thus, $-b$ is a lower bound for A. Suppose that $b' \leq a$ for all a in A. Then $-a \leq -b'$ for all a in A and so, $b \leq -b'$ since b is the least upper bound for $-A$. Thus, $b' \leq -b$ and hence, $-b$ is indeed the greatest lower bound for A. □

This lemma immediately gives the following theorem.

THEOREM 2.30. *Every nonempty set of real numbers that is bounded below has an infimum, that is, greatest lower bound.*

The consequences of the completeness property of the real numbers are quite profound as we now intend to demonstrate!

2.7.2. Existence of n-th roots. As a first consequence of the completeness property we show that any nonnegative real number has a unique nonnegative n-th root where $n \in \mathbb{N}$. In the following theorem we use the fact that if $\xi, \eta > 0$, then for any $k \in \mathbb{N}$,

$$1 < \xi \quad \Longrightarrow \quad \xi \leq \xi^k \qquad \text{and} \qquad \eta < 1 \quad \Longrightarrow \quad \eta^k \leq \eta.$$

These properties follow from the power rules in Theorem 2.22. E.g. $1 < \xi$ implies $1 = 1^{k-1} \leq \xi^{k-1}$ (with $=$ when $k = 1$ and with $<$ when $k > 1$); then multiplying $1 \leq \xi^{k-1}$ by ξ we get $\xi \leq \xi^k$. A similar argument shows that $\eta^k \leq \eta$.

THEOREM 2.31 (**Existence/uniqueness of n-th roots**). *Every nonnegative real number has a unique nonnegative n-th root.*

PROOF. First of all, uniqueness follows from the last power rule in Theorem 2.22. Note that the n-th root of zero exists and equals zero and certainly 1-th roots always exist. So, let $a > 0$ and $n \geq 2$; we shall prove that $\sqrt[n]{a}$ exists.

Step 1: We first define the tentative $\sqrt[n]{a}$ as a supremum. Let A be the set of real numbers x such that $x^n \leq a$. Certainly A contains 0, so A is nonempty. We claim that A is bounded above by $a+1$. To see this, observe that if $x \geq a+1$, then in particular $x > 1$, so for such x,

$$a < a + 1 \leq x \leq x^n \quad \Longrightarrow \quad x \notin A.$$

This shows that A is bounded above by $a+1$. Being nonempty and bounded above, by the axiom of completeness, A has a least upper bound, which we denote by $b \geq 0$. We shall prove that $b^n = a$, which proves our theorem. Well, either $b^n = a$, $b^n < a$, or $b^n > a$. We shall prove that the latter two cases cannot occur.

Step 2: Suppose that $b^n < a$. Let $0 < \varepsilon < 1$. Then $\varepsilon^m \le \varepsilon$ for any natural number m, so by the binomial theorem,

$$\begin{aligned}(b+\varepsilon)^n = \sum_{k=0}^{n} \binom{n}{k} b^k\, \varepsilon^{n-k} &= b^n + \sum_{k=0}^{n-1} \binom{n}{k} b^k\, \varepsilon^{n-k} \\ &\le b^n + \sum_{k=0}^{n-1} \binom{n}{k} b^k\, \varepsilon = b^n + \varepsilon c,\end{aligned}$$

where c is the positive number $c = \sum_{k=0}^{n-1} \binom{n}{k} b^k$. Since $b^n < a$, we have $(a - b^n)/c > 0$. Let ε equal $(a - b^n)/c$ or $1/2$, whichever is smaller (or equal to $1/2$ if $(a - b^n)/c = 1/2$). Then $0 < \varepsilon < 1$ and $\varepsilon \le (a - b^n)/c$, so

$$(b+\varepsilon)^n \le b^n + \varepsilon c \le b^n + \frac{a - b^n}{c} \cdot c = a.$$

This shows that $b + \varepsilon$ also belongs to A, which contradicts the fact that b is an upper bound for A.

Step 3: Now suppose that $b^n > a$. Then $b > 0$ (for if $b = 0$, then $b^n = 0 \not> a$). Given any $0 < \varepsilon < b$, we have $\varepsilon\, b^{-1} < 1$, which implies $-\varepsilon\, b^{-1} > -1$, so by Bernoulli's inequality (Theorem 2.7),

$$(b-\varepsilon)^n = b^n \left(1 - \varepsilon b^{-1}\right)^n \ge b^n \left(1 - n\varepsilon b^{-1}\right) = b^n - \varepsilon\, c,$$

where $c = nb^{n-1} > 0$. Since $a < b^n$, we have $(b^n - a)/c > 0$. Let ε equal $(b^n - a)/c$ or $b/2$, whichever is smaller (or equal to $b/2$ if $(b^n - a)/c = b/2$). Then $0 < \varepsilon < b$ and $\varepsilon \le (b^n - a)/c$, which implies that $-\varepsilon c \ge -(b^n - a)$. Therefore,

$$(b-\varepsilon)^n \ge b^n - \varepsilon c \ge b^n - (b^n - a) = a.$$

This shows that $b - \varepsilon$ is an upper bound for A, which contradicts the fact that b is the least upper bound for A.

□

In particular, $\sqrt{2}$ exists and, as we already know, is an irrational number. Here are proofs of the familiar root rules memorized from high school.

Theorem 2.32 (**Root rules**). *For any nonnegative real numbers a and b and natural number n, we have*

$$\sqrt[n]{ab} = \sqrt[n]{a}\,\sqrt[n]{b}, \qquad \sqrt[m]{\sqrt[n]{a}} = \sqrt[mn]{a}.$$

Moreover,

$$a < b \quad \Longleftrightarrow \quad \sqrt[n]{a} < \sqrt[n]{b}.$$

Proof. Let $x = \sqrt[n]{a}$ and $y = \sqrt[n]{b}$. Then, $x^n = a$ and $y^n = b$, so

$$(xy)^n = x^n\, y^n = ab.$$

By uniqueness of n-th roots, we must have $xy = \sqrt[n]{ab}$. This proves the first identity. The second identity is proved similarly. Finally, by our power rules theorem (Theorem 2.22), we have $\sqrt[n]{a} < \sqrt[n]{b} \Longleftrightarrow (\sqrt[n]{a})^n < \left(\sqrt[n]{b}\right)^n \Longleftrightarrow a < b$, which proves the last statement of our theorem. □

Another way to write these root rules are

$$(ab)^{\frac{1}{n}} = a^{\frac{1}{n}} b^{\frac{1}{n}}, \qquad (a^{\frac{1}{n}})^{\frac{1}{m}} = a^{\frac{1}{mn}}, \quad \text{and} \quad a < b \quad \Longleftrightarrow \quad a^{\frac{1}{n}} < b^{\frac{1}{n}}.$$

Given any $a \in \mathbb{R}$ with $a \geq 0$ and $r = m/n$ where $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, we define

$$a^r := \big(a^{1/n}\big)^m, \tag{2.27}$$

provided that $a \neq 0$ when $m < 0$. One can check that the right-hand side is defined independent of the representation of r; that is, if $r = p/q = m/n$ for some other $p \in \mathbb{Z}$ and $q \in \mathbb{N}$, then $(a^{1/q})^p = (a^{1/n})^m$. Combining the power rules theorem for integer powers and the root rules theorem above, we get

Theorem 2.33 (**Power rules for rational powers**). *For $a, b \in \mathbb{R}$ with $a, b \geq 0$, and $r, s \in \mathbb{Q}$, we have*

$$a^r \cdot a^s = a^{r+s}; \quad a^r \cdot b^r = (ab)^r; \quad (a^r)^s = a^{rs},$$

provided that the individual powers are defined (e.g. a and b are nonzero if an exponent is negative). If r is nonnegative and $a, b \geq 0$, then

$$a < b \quad \Longleftrightarrow \quad a^r < b^r.$$

We shall define a^x for any real number x in Section 4.6 and prove a similar theorem (see Theorem 4.32); see also Exercise 9 for another way to define a^x.

2.7.3. The Archimedean property and its consequences.

Another consequence of the completeness property is the following "obvious" fact.

Lemma 2.34. *$\mathbb{N}$ is not bounded above and $\mathbb{Z}$ is not bounded above nor below.*

Proof. We only prove the claim for $\mathbb{N}$ leaving the claim for $\mathbb{Z}$ to you. Assume, for sake of achieving a contradiction, that $\mathbb{N}$ is bounded above. Then the set $\mathbb{N}$ must have a least upper bound, say b. Since the number $b - 1$ is smaller than the least upper bound b, there must be a natural number m such that $b - 1 < m$, which implies that $b < m + 1$. However, $m + 1$ is a natural number, so b cannot be an upper bound for $\mathbb{N}$, a contradiction. □

This lemma yields many useful results.

Theorem 2.35 (**The $1/n$-principle**). *Given any real number $x > 0$, there is a natural number n such that $\frac{1}{n} < x$.*

Proof. Indeed, since $\mathbb{N}$ is not bounded above, $\frac{1}{x}$ is not an upper bound so there is an $n \in N$ such that $\frac{1}{x} < n$. This implies that $\frac{1}{n} < x$ and we're done. □

Here's an example showing the $1/n$-principle in action.

Example 2.25. Let

$$A = \left\{ 1 - \frac{3}{n} \,;\, n = 1, 2, 3, \ldots \right\}.$$

We shall prove that $\sup A = 1$ and $\inf A = -2$. (Please draw a few points of A on a number line to see why these values for the sup and inf are reasonable.)

To show that $\sup A = 1$ we need to prove two things: That 1 is an upper bound for A and that 1 is the least of all upper bounds for A. First, we show that 1 is an upper bound. To see this, observe that for any $n \in \mathbb{N}$,

$$\frac{3}{n} \geq 0 \quad \Longrightarrow \quad -\frac{3}{n} \leq 0 \quad \Longrightarrow \quad 1 - \frac{3}{n} \leq 1.$$

Thus, for all $a \in A$, $a \leq 1$, so 1 is indeed an upper bound for A. Second, we must show that 1 is the least of all upper bounds. So, assume that $c < 1$; we'll show that c cannot be an upper bound by showing that there is an $a \in A$ such that $c < a$; that is, there is an $n \in \mathbb{N}$ such that $c < 1 - 3/n$. Observe that

$$c < 1 - \frac{3}{n} \iff \frac{3}{n} < 1 - c \iff \frac{1}{n} < \frac{1-c}{3}. \tag{2.28}$$

Since $c < 1$, we have $(1-c)/3 > 0$, so by the $1/n$-principle, there exists an $n \in \mathbb{N}$ such that $1/n < (1-c)/3$. Hence, by (2.28), there is an $n \in \mathbb{N}$ such that $c < 1-3/n$. This shows that c is not an upper bound for A.

To show that $\inf A = -2$ we need to prove two things: That -2 is a lower bound and that -2 is the greatest of all lower bounds. First, to prove that -2 is a lower bound, observe that for any $n \in \mathbb{N}$,

$$\frac{3}{n} \leq 3 \implies -3 \leq -\frac{3}{n} \implies -2 \leq 1 - \frac{3}{n}.$$

Thus, for all $a \in A$, $-2 \leq a$, so -2 is indeed a lower bound for A. Second, to see that -2 is the greatest of all lower bounds, assume that $-2 < c$; we'll show that c cannot be a lower bound by showing there is an $a \in A$ such that $a < c$; that is, there is an $n \in \mathbb{N}$ such that $1 - 3/n < c$. In fact, simply take $n = 1$. Then $1 - 3/n = 1 - 3 = -2 < c$. This shows that c is not a lower bound for A.

Here's another useful consequence of the fact that $\mathbb{N}$ is not bounded above.

Theorem 2.36 (**Archimedean property**). [11] *Given a real number $x > 0$ and a real number y, there is a unique integer n such that*

$$nx \leq y < (n+1)x.$$

In particular, with $x = 1$, given any real number y there is a unique integer n such that $n \leq y < n+1$, a fact that is obvious from viewing the real numbers as a line.

Proof. Dividing $nx \leq y < (n+1)x$ by x, we need to prove that there is a unique integer n such that

$$n \leq z < n+1, \qquad \text{where } z = \frac{y}{x}.$$

We first prove existence existence. Since $\mathbb{N}$ is not bounded above, there is a $k \in \mathbb{N}$ such that $1 - z < k$, or adding z's, $1 < z + k$. Again using that $\mathbb{N}$ is not bounded above, the set $A = \{m \in \mathbb{N}\,;\, z + k < m\}$ is not empty, so by the well-ordering of $\mathbb{N}$, A contains a least element, say $\ell \in \mathbb{N}$. Then $z + k < \ell$ (because $\ell \in A$) and $\ell - 1 \leq z + k$ (because on the other hand, if $z + k < \ell - 1$, then setting $m = \ell - 1 < \ell$ and recalling that $1 < z + k$, we see that $m \in \mathbb{N}$ and $z + k < m$, so $m \in A$ is smaller than ℓ contradicting that ℓ is the least element of A). Thus,

$$\ell - 1 \leq z + k < \ell \implies n \leq z < n+1,$$

where $n = \ell - 1 - k \in \mathbb{Z}$.

To prove uniqueness, assume that $n \leq z < n+1$ and $m \leq z < m+1$ for $m, n \in \mathbb{Z}$. These inequalities imply that $n \leq z < m+1$, so $n < m+1$, and that $m \leq z < n+1$, so $m < n+1$. Thus, $n < m+1 < (n+1)+1 = n+2$, or $0 < m - n + 1 < 2$. This implies that $m - n + 1 = 1$, or $m = n$. □

[11] The "Archimedean property" might equally well be called the "Eudoxus property" after Eudoxus of Cnidus (408 B.C.–355 B.C.); see [**169**] and [**120**, p. 7].

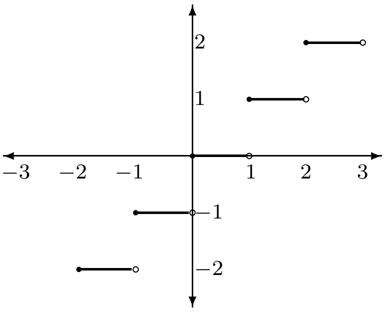


FIGURE 2.6. The greatest integer function.

We remark that some authors replace the integer n by the integer $n-1$ in the Archimedean property so it reads: Given a real number $x > 0$ and a real number y, there is a unique integer n such that $(n-1)x \leq y < nx$. We'll use this formulation of the Archimedean property in the proof of Theorem 2.37 below.

Using the Archimedean property, we can define the **greatest integer function** as follows (see Figure 2.6): Given any $a \in \mathbb{R}$, we define $\lfloor a \rfloor$ as the greatest integer less than or equal to a, that is, $\lfloor a \rfloor$ is the unique integer n satisfying the inequalities $n \leq a < n+1$. This function will come up various times in the sequel. We now prove an important fact concerning the rational and irrational numbers.

THEOREM 2.37 (**Density of the (ir)rationals**). *Between any two real numbers is a rational and irrational number.*

PROOF. Let $x < y$. We first prove that there is a rational number between x and y. Indeed, $y - x > 0$, so by our corollary there is a natural number m such that $1/m < y - x$. By the Archimedean principle, there is an integer n such that

$$n - 1 \leq mx < n \quad \Longrightarrow \quad \frac{n}{m} - \frac{1}{m} \leq x < \frac{n}{m}.$$

In particular, $x < n/m$, and

$$\frac{n}{m} \leq \frac{1}{m} + x < (y - x) + x = y \quad \Longrightarrow \quad \frac{n}{m} < y.$$

Thus, the rational number n/m is between x and y.

To prove that between x and y is an irrational number, note that $x - \sqrt{2} < y - \sqrt{2}$, so by what we just proved above, there is a rational number r such that $x - \sqrt{2} < r < y - \sqrt{2}$. Adding $\sqrt{2}$, we obtain

$$x < \xi < y, \qquad \text{where } \ \xi = r + \sqrt{2}.$$

Note that ξ is irrational, for if it were rational, then $\sqrt{2} = \xi - r$ would also be rational, which we know is false. This completes our proof. □

2.7.4. The nested intervals property. A sequence of sets $\{A_n\}$ is said to be **nested** if

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \supseteq A_n \supseteq A_{n+1} \supseteq \cdots ,$$

that is $A_k \supseteq A_{k+1}$ for each k.

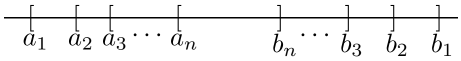


FIGURE 2.7. Nested Intervals.

Example 2.26. If $A_n = \left(0, \frac{1}{n}\right)$, then $\{A_n\}$ is a nested sequence. Note that

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right) = \varnothing.$$

Indeed, if $x \in \bigcap A_n$, which means that $x \in \left(0, \frac{1}{n}\right)$ for every $n \in \mathbb{N}$, then $0 < x < 1/n$ for all $n \in \mathbb{N}$. However, by Theorem 2.35, there is an n such that $0 < 1/n < x$. This shows that $x \notin (0, 1/n)$, contradicting that $x \in \left(0, \frac{1}{n}\right)$ for every $n \in \mathbb{N}$. Therefore, $\bigcap A_n$ must be empty.

Example 2.27. Now on the other hand, if $A_n = \left[0, \frac{1}{n}\right]$, then $\{A_n\}$ is a nested sequence, but in this case,

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left[0, \frac{1}{n}\right] = \{0\} \neq \varnothing.$$

The difference between the first example and the second is that the second example is a nested sequence of closed and bounded intervals. Here, bounded means bounded above and below. It is a general fact that the intersection of a nested sequence of nonempty closed and bounded intervals is nonempty. This is the content of the nested intervals theorem.

THEOREM 2.38 (**Nested intervals theorem**). *The intersection of a nested sequence of nonempty closed and bounded intervals in $\mathbb{R}$ is nonempty.*

PROOF. Let $\{I_n = [a_n, b_n]\}$ be a sequence of nonempty closed and bounded intervals. Since we are given that this sequence is nested, we in particular have $I_2 = [a_2, b_2] \subseteq [a_1, b_1] = I_1$, and so $a_1 \le a_2 \le b_2 \le b_1$. Since $I_3 = [a_3, b_3] \subseteq [a_2, b_2]$, we have $a_2 \le a_3 \le b_3 \le b_2$, and so $a_1 \le a_2 \le a_3 \le b_3 \le b_2 \le b_1$. In general, we see that for any n,

$$a_1 \le a_2 \le a_3 \le \cdots \le a_n \le b_n \le \cdots \le b_3 \le b_2 \le b_1.$$

See Figure 2.7. Let $a = \sup\{a_k \,;\, k = 1, 2, \ldots\}$. Since $a_1 \le a_2 \le a_3 \le \cdots$, by definition of supremum $a_n \le a$ for each n. Also, since any b_n is an upper bound for the set $\{a_k \,;\, k = 1, 2, \ldots\}$, by definition of supremum, $a \le b_n$ for each n. Thus, $a \in I_n$ for each n, and our proof is complete. □

Example 2.28. The "bounded" assumption cannot be dropped, for if $A_n = [n, \infty)$, then $\{A_n\}$ is a nested sequence, but

$$\bigcap_{n=1}^{\infty} A_n = \varnothing.$$

We end this section with a discussion of maximums and minimums. Given any set A of real numbers, a number a is called the **maximum** of A if $a \in A$ and $a = \sup A$, in which case we write $a = \max A$. Similarly, a is called the **minimum** of A if $a \in A$ and $a = \inf A$, in which case we write $a = \min A$. For instance, $1 = \max(0, 1]$, but $(0, 1)$ has no maximum, only a supremum, which is also 1. In Problem 4, we prove that any finite set has a maximum.

Exercises 2.7.

1. What are the supremums and infimums of the following sets? Give careful proofs of your answers. The "$1/n$-principle" might be helpful in some of your proofs.

(a) $A = \{1 + \frac{5}{n}\,;\, n = 1, 2, 3, \ldots\}$ (b) $B = \{3 - \frac{8}{n^3}\,;\, n = 1, 2, 3, \ldots\}$

(c) $C = \big\{1 + (-1)^n \frac{1}{n}\,;\, n = 1, 2, 3, \ldots\big\}$ (d) $D = \big\{(-1)^n + \frac{1}{n}\,;\, n = 1, 2, 3, \ldots\big\}$

(e) $E = \big\{\sum_{k=1}^{n} \frac{1}{2^k}\,;\, n = 1, 2, \ldots\big\}$ (f) $F = \Big\{(-1)^n + \frac{(-1)^{n+1}}{n}\,;\, n = 1, 2, 3, \ldots\Big\}$.

2. Are the following sets bounded above? Are they bounded below? If the supremum or infimum exists, find it and prove your answer.

$$(a)\ A = \Big\{1 + n^{(-1)^n}\,;\, n = 1, 2, 3, \ldots\Big\} \quad , \quad (b)\ B = \Big\{2^{n(-1)^n}\,;\, n = 1, 2, 3, \ldots\Big\}.$$

3. (Various properties of supremums/infimums)
   (a) If $A \subseteq \mathbb{R}$ is bounded above and A contains one of its upper bounds, prove that this upper bound is in fact the supremum of A.
   (b) Let $A \subseteq \mathbb{R}$ be a nonempty bounded set. For $x, y \in \mathbb{R}$, define a new set $xA + y$ by $xA + y := \{xa + y\,;\, a \in A\}$. Consider the case $y = 0$. Prove that
$$x > 0 \implies \inf(xA) = x \inf A,\ \sup(xA) = x \sup A;$$
$$x < 0 \implies \inf(xA) = x \sup A,\ \inf(xA) = x \sup A.$$
   (c) With $x = 1$, prove that $\inf(A + y) = \inf(A) + y$ and $\sup(A + y) = \sup(A) + y$.
   (d) What are the formulas for $\inf(xA + y)$ and $\sup(xA + y)$?
   (e) If $A \subseteq B$ and B is bounded, prove that $\sup A \leq \sup B$ and $\inf B \leq \inf A$.
4. In this problem we prove some facts concerning maximums and minimums.
   (a) Let $A \subseteq \mathbb{R}$ be nonempty. An element $a \in A$ is called the **maximum**, respectively **minimum**, element of A if $a \geq x$, respectively $a \leq x$, for all $x \in A$. Prove A has a maximum (resp. minimum) if and only if $\sup A$ exists and $\sup A \in A$ (resp. $\inf A$ exists and $\inf A \in A$).
   (b) Let $A \subseteq \mathbb{R}$ and suppose that A has a maximum, say $a = \max A$. Given any $b \in \mathbb{R}$, prove that $A \cup \{b\}$ also has a maximum, and $\max(A \cup \{b\}) = \max\{a, b\}$.
   (c) Prove that a nonempty finite set of real numbers has a maximum and minimum, where by finite we mean a set of the form $\{a_1, a_2, \ldots, a_n\}$ where $a_1, \ldots, a_n \in \mathbb{R}$.
5. If $A \subseteq \mathbb{R}^+$ is nonempty and closed under addition, prove that A is not bounded above. (As a corollary, we get another proof that $\mathbb{N}$ is not bounded above.) If $A \subseteq (1, \infty)$ is nonempty and closed under multiplication, prove that A is not bounded above.
6. Using the Archimedean property, prove that if $a, b \in \mathbb{R}$ and $b - a > 1$, then there is an $n \in \mathbb{Z}$ such that $a < n < b$. Using this result can you give another proof that between any two real numbers there is a rational number?
7. If $a \in \mathbb{R}$, prove that $|a| = \sqrt{a^2}$.
8. Here are some more power rules for you to prove. Let $p, q \in \mathbb{Q}$.
   (a) If $p < q$ and $a > 1$, then $a^p < a^q$.
   (b) If $p < q$ and $0 < a < 1$, then $a^q < a^p$.
   (c) Let $a > 0$ and let $p < q$. Prove that $a > 1$ if and only if $a^p < a^q$.
9. (**Real numbers to real powers**) We define $0^x := 0$ for all $x > 0$; otherwise 0^x is undefined. We now define a^x for $a > 0$ and $x \in \mathbb{R}$. First, assume that $a \geq 1$ and $x \geq 0$.
   (a) Prove that $A = \{a^r\,;\, 0 \leq r \leq x\}$ is bounded above, where a^r is defined in (2.27). Define $a^x := \sup A$. Prove that if $x \in \mathbb{Q}$, then this definition of a^x agrees with the definition (2.27).
   (b) For $a, b, x, y \in \mathbb{R}$ with $a, b \geq 1$ and $x, y \geq 0$, prove that
$$a^x \cdot a^y = a^{x+y}; \quad a^x \cdot b^x = (ab)^x; \quad (a^x)^y = a^{xy}. \tag{2.29}$$
   (In the equality $(a^x)^y = a^{xy}$, you should first show that $a^x \geq 1$ so $(a^x)^y$ is defined.)

(c) If $0 < a < 1$ and $x \geq 0$, define $a^x := 1/(1/a)^x$; note that $1/a > 1$ so $(1/a)^x$ is defined. Finally, if $a > 0$ and $x < 0$, define $a^x := (1/a)^{-x}$; note that $-x > 0$ so $(1/a)^{-x}$ is defined. Prove (2.29) for any $a, b, x, y \in \mathbb{R}$ with $a, b > 0$ and $x, y \in \mathbb{R}$.

10. Let $p(x) = ax^2 + bx + c$ be a quadratic polynomial with real coefficients and with $a \neq 0$. Prove that $p(x)$ has a real root (that is, an $x \in \mathbb{R}$ with $p(x) = 0$) if and only if $b^2 - 4ac \geq 0$, in which case, the root(s) are given by the quadratic formula:
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$
11. Let $\{I_n = [a_n, b_n]\}$ be a nested sequence of nonempty closed and bounded intervals and put $A = \{a_n \,;\, n \in \mathbb{N}$ and $B = \{b_n \,;\, n \in \mathbb{N}\}$. Show that $\sup A$ and $\inf B$ exist and $\bigcap I_n = [\sup A, \inf B]$.
12. In this problem we give a characterization of the completeness axiom (**C**) of $\mathbb{R}$ in terms of intervals as explained by Christian [**50**]. A subset A of $\mathbb{R}$ is **convex** if given any x and y in A and $t \in \mathbb{R}$ with $x < t < y$, we have $t \in A$.
    (a) Assume axiom (**C**). Prove that all convex subsets of $\mathbb{R}$ are intervals.
    (b) Assume that all convex subsets of $\mathbb{R}$ are intervals. Prove the completeness property (**C**) of $\mathbb{R}$. Suggestion: Let I be the set of all upper bounds of a nonempty set A that is bounded above. Show that I is convex.

    This problem shows that the completeness axiom is equivalent to the statement that all convex sets are intervals.

2.8. m-dimensional Euclidean space

The plane $\mathbb{R}^2$ is said to be two-dimensional because to locate a point in the plane requires two points, its ordered pair of coordinates. Similarly, we are all familiar with $\mathbb{R}^3$, which is said to be three-dimensional because to represent any point in space we need an ordered triple of real numbers. In this section we generalize these considerations to m-dimensional space $\mathbb{R}^m$ and study its properties.

2.8.1. The vector space structure of $\mathbb{R}^m$. Recall that the set $\mathbb{R}^m$ is just the product $\mathbb{R}^m := \mathbb{R} \times \cdots \times \mathbb{R}$ (m copies of $\mathbb{R}$), or explicitly, the set of all m-tuples of real numbers,
$$\mathbb{R}^m := \big\{(x_1, \ldots, x_m) \,;\, x_1, \ldots, x_m \in \mathbb{R}\big\}.$$
We call elements of $\mathbb{R}^m$ **vectors** (or **points**) and we use the notation 0 for the m-tuple of zeros $(0, \ldots, 0)$ (m zeros); it will always be clear from context whether 0 refers to the real number zero or the m-tuple of zeros. In elementary calculus, we usually focus on the case when $m = 2$ or $m = 3$; e.g. when $m = 2$,
$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \big\{(x_1, x_2) \,;\, x_1, x_2 \in \mathbb{R}\big\},$$
and in this case, the zero vector "0" $= (0, 0)$.

Given any $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ in $\mathbb{R}^m$ and real number a, we define
$$x + y := (x_1 + y_1, \ldots, x_m + y_m) \qquad \text{and} \qquad a\,x := (ax_1, \ldots, ax_m).$$
We also define
$$-x := (-x_1, \ldots, -x_m).$$
With these definitions, observe that
$$x + y = (x_1 + y_1, \ldots, x_m + y_m) = (y_1 + x_1, \ldots, y_m + x_m) = y + x,$$
and
$$x + 0 = (x_1 + 0, \ldots, x_m + 0) = (x_1, \ldots, x_m) = x$$

and similarly, $0+x = x$. These computations prove properties (A1) and (A3) below, and you can check that the following further properties of addition are satisfied: Addition satisfies

(A1) $x + y = y + x$; (commutative law)
(A2) $(x + y) + z = x + (y + z)$; (associative law)
(A3) there is an element 0 such that $x + 0 = x = 0 + x$; (additive identity)
(A4) for each x there is a $-x$ such that

$$x + (-x) = 0 \quad \text{and} \quad (-x) + x = 0. \quad \text{(additive inverse)}$$

Of course, we usually write $x + (-y)$ as $x - y$.

Multiplication by real numbers satisfies

(M1) $1 \cdot x = x$; (multiplicative identity)
(M2) $(a\, b)\, x = a\, (bx)$; (associative law)

and finally, addition and multiplication are related by

(D) $a(x + y) = ax + ay$ and $(a + b)x = ax + bx$. (distributive law)

We remark that any set, say with elements denoted by $x, y, z, \ldots$, called **vectors**, with an operation of "+" and an operation of multiplication by real numbers that satisfy properties (**A1**) – (**A4**), (**M1**) – (**M2**), and (**D**), is called a **real vector space**. If the scalars $a, b, 1$ in (**M1**) – (**M2**) and (**D**) are elements of a field $\mathbb{F}$, then we say that the vector space is an $\mathbb{F}$ vector space or a vector space **over** $\mathbb{F}$. In particular, $\mathbb{R}^m$ is a real vector space.

2.8.2. Inner products. We now review inner products, also called dot products in elementary calculus. We all probably know that given any two vectors $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ in $\mathbb{R}^3$, the dot product $x \cdot y$ is the number

$$x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

We generalize this to $\mathbb{R}^m$ as follows: If $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$, then we define the **inner product** (also called the **dot product** or **scalar product**) $\langle x, y\rangle$ as the real number

$$\boxed{\langle x, y\rangle := x_1 y_1 + x_2 y_2 + \cdots + x_m y_m = \sum_{j=1}^{m} x_j y_j.}$$

It is also common to denote $\langle x, y\rangle$ by $x \cdot y$ or (x, y), but we prefer the angle bracket notation $\langle x, y\rangle$, which is popular in physics, because the dot "$\cdot$" can be confused with multiplication and the parentheses "(,)" can be confused with ordered pair.

In the following theorem we summarize some of the main properties of $\langle \cdot, \cdot\rangle$.

Theorem 2.39. *For any vectors x, y, z in $\mathbb{R}^m$ and real number a,*

(i) $\langle x, x\rangle \geq 0$ *and* $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
(ii) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ *and* $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
(iii) $\langle a\, x, y\rangle = a\langle x, y\rangle$ *and* $\langle x, a\, y\rangle = a\langle x, y\rangle$.
(iv) $\langle x, y\rangle = \langle y, x\rangle$.

Proof. To prove *(i)*, just note that

$$\langle x, x\rangle = x_1^2 + x_2^2 + \cdots + x_m^2$$

and $x_j^2 \geq 0$ for each j. If $\langle x, x\rangle = 0$, then as the only way a sum of nonnegative numbers is zero is that each number is zero, we must have $x_j^2 = 0$ for each j. Hence, every $x_j = 0$ and therefore, $x = (x_1, \ldots, x_m) = 0$. Conversely, if $x = 0$, that is, $x_1 = 0, \ldots, x_m = 0$, then of course $\langle x, x \rangle = 0$ too. This concludes the proof of *(i)*.

To prove *(ii)*, we just compute:

$$\langle x + y, z \rangle = \sum_{j=1}^{m} (x_j + y_j)\, z_j = \sum_{j=1}^{m} (x_j z_j + y_j z_j) = \langle x, z \rangle + \langle y, z \rangle.$$

The other identity $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ is proved similarly. The proofs of *(iii)* and *(iv)* are also simple computations, so we leave their proofs to the reader. □

We remark that any real vector space V with an operation that assigns to every two vectors x and y in V a real number $\langle x, y \rangle$ satisfying properties *(i)* – *(iv)* of Theorem 2.39 is called a **real inner product space** and the operation $\langle \cdot, \cdot \rangle$ is called an **inner product** on V. In particular, $\mathbb{R}^m$ is a real inner product space.

2.8.3. The norm in $\mathbb{R}^m$. Recall that the length of a vector $x = (x_1, x_2, x_3)$ in $\mathbb{R}^3$ is just the distance of the point x from the origin, or

$$|x| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

We generalize these considerations as follows. The **length** or **norm** of a vector $x = (x_1, \ldots, x_m)$ in $\mathbb{R}^m$ is by definition the nonnegative real number

$$\boxed{|x| := \sqrt{x_1^2 + \cdots + x_m^2} = \sqrt{\langle x, x \rangle} \geq 0.}$$

We interpret the norm $|x|$ as the length of the vector x, or the distance of x from the origin 0. In particular, the squared length $|x|^2$ of the vector x is given by

$$|x|^2 = \langle x, x \rangle.$$

Warning: For $m > 1$, $|x|$ does not mean absolute value of a real *number* x, it means norm of a *vector* x. However, if $m = 1$, then "norm" and "absolute value" are the same, because for $x = x_1 \in \mathbb{R}^1 = \mathbb{R}$, the above definition of norm is $\sqrt{x_1^2}$, which is exactly the absolute value of x_1 according to Problem 7 in Exercises 2.7.

The following inequality relates the norm and the inner product. It is commonly called the **Schwarz inequality** or **Cauchy-Schwarz inequality** after Hermann Schwarz (1843–1921) who stated it for integrals in 1885 and Augustin Cauchy (1789–1857) who stated it for sums in 1821. However, (see [**94**] for the history) it perhaps should be called the **Cauchy-Bunyakovskiĭ-Schwarz inequality** because Viktor Bunyakovskiĭ (1804–1889), a student of Cauchy, published a related inequality 25 years before Schwarz. (Note: There is no "t" before the "z.")

THEOREM 2.40 (**Schwarz inequality**). *For any vectors x, y in $\mathbb{R}^m$, we have*

$$\boxed{|\langle x, y \rangle| \leq |x|\,|y| \qquad \textbf{\textit{Schwarz inequality}}.}$$

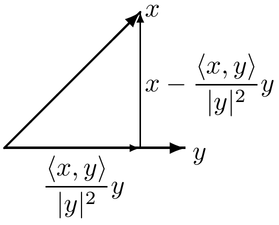

x
$x - \frac{\langle x, y\rangle}{|y|^2} y$
y
$\frac{\langle x, y\rangle}{|y|^2} y$

FIGURE 2.8. The projection of x onto y is $\frac{\langle x,y\rangle}{|y|^2} y$ and the projection of x onto the orthogonal complement of y is $x - \frac{\langle x,y\rangle}{|y|^2} y$.

PROOF. If $y = 0$, then both sides of Schwarz's inequality are zero, so we may assume that $y \neq 0$. Taking the squared length of the vector $x - \frac{\langle x,y\rangle}{|y|^2} y$, we get

$$\begin{aligned} 0 \leq \left| x - \frac{\langle x, y\rangle}{|y|^2} y \right|^2 &= \left\langle x - \frac{\langle x, y\rangle}{|y|^2}\, y, x - \frac{\langle x, y\rangle}{|y|^2}\, y \right\rangle \\ &= \langle x, x\rangle - \frac{\langle x, y\rangle}{|y|^2} \langle x, y\rangle - \frac{\langle x, y\rangle}{|y|^2} \langle y, x\rangle + \frac{\langle x, y\rangle \langle x, y\rangle}{|y|^4} \langle y, y\rangle \\ &= |x|^2 - \frac{\langle x, y\rangle^2}{|y|^2} - \frac{\langle x, y\rangle^2}{|y|^2} + \frac{\langle x, y\rangle^2}{|y|^2}. \end{aligned}$$

Cancelling the last two terms, we see that

$$0 \leq |x|^2 - \frac{|\langle x, y\rangle|^2}{|y|^2} \quad \Longrightarrow \quad |\langle x, y\rangle|^2 \leq |x|^2 |y|^2.$$

Taking square roots proves the Schwarz inequality. As a side remark (skip this if you're not interested), the vector $x - \frac{\langle x,y\rangle}{|y|^2} y$ that we took the squared length of didn't come out of a hat. Recall from your "multi-variable calculus" or "vector calculus" course, that the projection of x onto y and the projection of x onto the orthogonal complement of y are given by $\frac{\langle x,y\rangle}{|y|^2} y$ and $x - \frac{\langle x,y\rangle}{|y|^2}\, y$, respectively, see Figure 2.8. Thus, all we did above was take the squared length of the projection of x onto the orthogonal complement of y. □

In the following theorem, we list some of the main properties of the norm $|\cdot|$.

THEOREM 2.41. *For any vectors x, y in $\mathbb{R}^m$ and real number a,*

(i) $|x| \geq 0$ *and* $|x| = 0$ *if and only if* $x = 0$.
(ii) $|a\, x| = |a|\, |x|$.
(iii) $|x + y| \leq |x| + |y|$ ***(triangle inequality)****.*
(iv) $|\,|x| - |y|\,| \leq |x \pm y| \leq |x| + |y|$.

PROOF. *(i)* follows from Property *(i)* of Theorem 2.39. To prove *(ii)*, observe that

$$|ax| = \sqrt{\langle ax, ax\rangle} = \sqrt{a^2 \langle x, x\rangle} = |a| \sqrt{\langle x, x\rangle} = |a|\, |x|,$$

and therefore $|ax| = |a|\,|x|$. To prove the triangle inequality, we use the Schwarz inequality to get

$$\begin{aligned} |x+y|^2 = \langle x+y, x+y\rangle &= |x|^2 + \langle x,y\rangle + \langle y,x\rangle + |y|^2 \\ &= |x|^2 + 2\langle x,y\rangle + |y|^2 \\ &\le |x|^2 + 2|\langle x,y\rangle| + |y|^2 \\ &\le |x|^2 + 2|x|\,|y| + |y|^2, \end{aligned}$$

where we used the Schwarz inequality at the last step. Thus,

$$|x+y|^2 \le (|x| + |y|)^2.$$

Taking the square root of both sides proves the triangle inequality.

The second half of *(iv)* follows from the triangle inequality:

$$|x \pm y| = |x + (\pm 1)y| \le |x| + |(\pm 1)y| = |x| + |y|.$$

To prove the first half $|\,|x| - |y|\,| \le |x \pm y|$ we use the triangle inequality to get

$$\begin{aligned} |x| - |y| = |(x-y) + y| - |y| \le |x-y| &+ |y| - |y| = |x-y| \\ \Longrightarrow \quad & |x| - |y| \le |x-y|. \end{aligned} \tag{2.30}$$

Switching the letters x and y in (2.30), we get $|y| - |x| \le |y - x|$ or $-(|x| - |y|) \le |x - y|$. Combining this with (2.30), we see that

$$|x| - |y| \le |x-y| \quad \text{and} \quad -(|x| - |y|) \le |x-y| \quad \Longrightarrow \quad |\,|x| - |y|\,| \le |x-y|,$$

where we used the definition of absolute value of the real number $|x| - |y|$. Replacing y with $-y$ and using that $|-y| = |y|$, we get $|\,|x| - |y|\,| \le |x+y|$. This finishes the proof of *(iv)*. □

We remark that any real vector space V with an operation that assigns to every vector x in V a nonnegative real number $|x|$, such that $|\cdot|$ satisfies properties *(i)* – *(iii)* of Theorem 2.41 is called a **real normed space** and the operation $|\cdot|$ is called a **norm** on V. In particular, $\mathbb{R}^m$ is a real normed space. The exercises explore different norms on $\mathbb{R}^m$.

In analogy with the distance between two real numbers, we define the **distance** between two vectors x and y in $\mathbb{R}^m$ to be the number

$$|x - y|.$$

In particular, the triangle inequality implies that given any other vector z, we have

$$|x-y| = |(x-z) + (z-y)| \le |x-z| + |z-y|,$$

that is,

$$|x-y| \le |x-z| + |z-y|. \tag{2.31}$$

This inequality is the "genuine" triangle inequality since it represents the geometrically intuitive fact that the distance between two points x and y is shorter than the distance transversed by going from x to z and then from z to y; see Figure 2.9.

Finally, we remark that the norm $|\cdot|$ on $\mathbb{R}^m$ is sometimes called the **ball norm** on $\mathbb{R}^m$ for the following reason. Let $r > 0$ and take $m = 3$. Then $|x| < r$ means that $\sqrt{x_1^2 + x_2^2 + x_3^2} < r$, or squaring both sides, we get

$$x_1^2 + x_2^2 + x_3^2 < r^2,$$

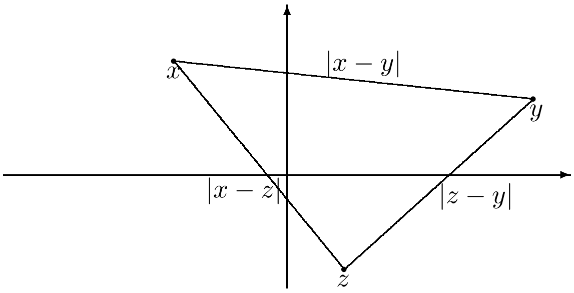


FIGURE 2.9. Why the "genuine" triangle inequality (2.31) is obvious.

which simply says that x is inside the ball of radius r. So, if $c \in \mathbb{R}^3$, then $|x-c| < r$ just means that

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 < r^2,$$

which is to say, x is inside the ball of radius r that is centered at the point $c = (c_1, c_2, c_3)$. Generalizing this notion to m-dimensional space, given c in $\mathbb{R}^m$, we call the set of all x such that $|x - c| < r$, or after squaring both sides,

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 + \cdots + (x_m - c_m)^2 < r^2,$$

the **open ball** of radius r centered at c. We denote this set by B_r, or $B_r(c)$ to emphasize that the center of the ball is c. Therefore,

$$\boxed{B_r(c) := \{x \in \mathbb{R}^m \,;\, |x - c| < r\}.} \tag{2.32}$$

The set of x with $<$ replaced by $\leq$ is called the **closed ball** of radius r centered at c and is denoted by $\overline{B}_r$ or $\overline{B}_r(c)$,

$$\boxed{\overline{B}_r(c) := \{x \in \mathbb{R}^m \,;\, |x - c| \leq r\}.}$$

If $m = 1$, then the ball concept reduces to intervals in $\mathbb{R}^1 = \mathbb{R}$:

$$\begin{aligned} x \in B_r(c) \quad &\Longleftrightarrow \quad |x-c| < r \quad \Longleftrightarrow \quad -r < x - c < r \\ &\Longleftrightarrow \quad c - r < x < c + r \quad \Longleftrightarrow \quad x \in (c - r, c + r). \end{aligned}$$

Thus, for $m = 1$, $B_r(c)$ is just the open interval centered at c of length $2r$. For $m = 1$, $\overline{B}_r(c)$ is just the closed interval centered at c of length $2r$.

EXERCISES 2.8.

1. Let $x, y \in \mathbb{R}^m$. Prove that
$$|x+y|^2 + |x-y|^2 = 2|x|^2 + 2|y|^2 \quad \textbf{(parallelogram law)}.$$
Vectors x and y are said to be **orthogonal** if $\langle x, y \rangle = 0$. Prove that x and y are orthogonal if and only if
$$|x+y|^2 = |x|^2 + |y|^2 \quad \textbf{(Pythagorean theorem)}.$$
2. (**Schwarz's inequality, Proof II**) Here's another way to prove Schwarz's inequality.
(a) For any real numbers a and b, prove that
$$\boxed{ab \leq \frac{1}{2}(a^2 + b^2).}$$
(b) Let $x, y \in \mathbb{R}^m$ with $|x| = 1$ and $|y| = 1$. Using (a), prove that $|\langle x, y \rangle| \leq 1$.

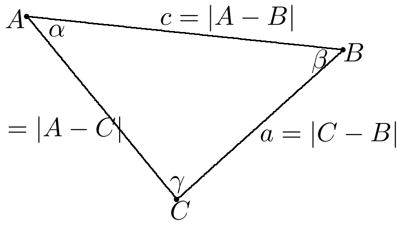


FIGURE 2.10. Triangle for the law of cosines.

(c) Now let x and y be arbitrary nonzero vectors of $\mathbb{R}^m$. Applying (b) to the vectors $x/|x|$ and $y/|y|$, derive Schwarz's inequality.

3. (**Schwarz's inequality, Proof III**) Here's an "algebraic" proof. Let $x, y \in \mathbb{R}^m$ with $y \neq 0$ and let $p(t) = |x + ty|^2$ for $t \in \mathbb{R}$. Note that $p(t) \geq 0$ for all t.
(a) Show that $p(t)$ can be written in the form $p(t) = a\,t^2 + 2b\,t + c$ where a, b, c are real numbers with $a \neq 0$.
(b) Using the fact that $p(t) \geq 0$ for all t, prove the Schwarz inequality. Suggestion: Write $p(t) = a(t + b/a)^2 + (c - b^2/a)$.

4. Prove that for any vectors x and y in $\mathbb{R}^m$, we have

$$2|x|^2|y|^2 - 2\left(\sum_{n=1}^m x_n\,y_n\right)^2 = \sum_{k=1}^m\sum_{\ell=1}^m (x_k y_\ell - x_\ell y_k)^2 \quad (\textbf{Lagrange identity}).$$

after Joseph-Louis Lagrange (1736–1813). Suggestion: Show that

$$\begin{aligned} 2|x|^2|y|^2 - 2\left(\sum_{n=1}^m x_n\,y_n\right)^2 &= 2\left(\sum_{k=1}^m x_k^2\right)\left(\sum_{\ell=1}^m y_\ell^2\right) - 2\left(\sum_{k=1}^m x_k\,y_k\right)\left(\sum_{\ell=1}^m x_\ell\,y_\ell\right) \\ &= 2\sum_{k=1}^m\sum_{\ell=1}^m x_k^2\,y_\ell^2 - 2\sum_{k=1}^m\sum_{\ell=1}^m x_k\,y_k\,x_\ell\,y_\ell \end{aligned}$$

and prove that $\sum_{k=1}^m \sum_{\ell=1}^m (x_k y_\ell - x_\ell y_k)^2$, when expanded out, has the same form.

5. (**Schwarz's inequality, Proof IV**)
(a) Prove the Schwarz inequality from Lagrange's identity.
(b) Using Lagrange's identity, prove that equality holds in the Schwarz inequality (that is, $|\sum_{n=1}^m x_n\,y_n| = |x|\,|y|$) if and only if x and y are collinear, which is to say, $x = 0$ or $y = c\,x$ for some $c \in \mathbb{R}$.
(c) Now show that equality holds in the triangle inequality (that is, $|x+y| = |x| + |y|$) if and only if $x = 0$ or $y = c\,x$ for some $c \geq 0$.

6. (**Laws of trigonometry**) In this problem we assume knowledge of the trigonometric functions; see Section 4.7 for a rigorous development of these functions. By the Schwarz inequality, given any two nonzero vectors $x, y \in \mathbb{R}^m$, we have $\frac{|\langle x,y\rangle|}{|x|\,|y|} \leq 1$. In particular, there is a unique angle $\theta \in [0, \pi]$ such that $\cos\theta = \frac{\langle x,y\rangle}{|x|\,|y|}$. The number θ is by definition the **angle** between the vectors x, y.
(a) Consider the triangle labelled as in Figure 2.10. Prove the following:

$$\begin{aligned} a^2 &= b^2 + c^2 - 2bc\cos\alpha \\ b^2 &= a^2 + c^2 - 2ac\cos\beta \qquad (\textbf{Law of cosines}) \\ c^2 &= a^2 + b^2 - 2ab\cos\gamma. \end{aligned}$$

Suggestion: To prove the last equality, observe that $c^2 = |A - B|^2 = |x - y|^2$ where $x = A - C$, $y = B - C$. Compute the dot product $|x - y|^2 = \langle x - y, x - y\rangle$.

(b) Using that $\sin^2\alpha = 1 - \cos^2\alpha$ and that $a^2 = b^2 + c^2 - 2bc\cos\alpha$, prove that

$$\frac{\sin^2\alpha}{a^2} = 4\frac{s(s-a)(s-b)(s-c)}{a^2\,b^2\,c^2}, \tag{2.33}$$

where $s := (a+b+c)/2$ is called the **semiperimeter** . From (2.33), conclude that

$$\frac{\sin\alpha}{a} = \frac{\sin\beta}{b} = \frac{\sin\gamma}{c} \qquad (\textbf{Law of sines}).$$

(c) Assume the formula: Area of a triangle $= \frac{1}{2}$ base × height. Use this formula together with (2.33) to prove that the area of the triangle in Figure 2.10 is

$$\text{Area} = \sqrt{s(s-a)(s-b)(s-c)}, \qquad (\textbf{Heron's formula}),$$

a formula named after Heron of Alexandria (10–75). .

7. Given any $x = (x_1, x_2, \ldots, x_m)$ in $\mathbb{R}^m$, define

$$\|x\|_\infty := \max\bigl\{|x_1|, |x_2|, \ldots, |x_m|\bigr\},$$

called the **sup (or supremum) norm** . . . of course, we need to show this is a norm.

(a) Show that $\|\cdot\|_\infty$ defines a norm on $\mathbb{R}^m$, that is, $\|\cdot\|_\infty$ satisfies properties *(i) – (iii)* of Theorem 2.41.

(b) In $\mathbb{R}^2$, what are the set of all points $x = (x_1, x_2)$ such that $\|x\|_\infty \le 1$? Draw a picture of this set. Do you see why $\|\cdot\|_\infty$ is sometimes called the **box norm**?

(c) Show that for any x in $\mathbb{R}^m$,

$$\|x\|_\infty \le |x| \le \sqrt{m}\,\|x\|_\infty,$$

where $|x|$ denotes the usual "ball norm" of x.

(d) Let $\overline{B}_r$ denote the closed ball in $\mathbb{R}^m$ of radius r centered at the origin (the set of points x in $\mathbb{R}^m$ such that $|x| \le r$). Let $\overline{\text{Box}}_r$ denote the closed ball in $\mathbb{R}^m$ of radius r in the box norm centered at the origin (the set of points x in $\mathbb{R}^m$ such that $\|x\|_\infty \le r$). Show that

$$\overline{B}_1 \subseteq \overline{\text{Box}}_1 \subseteq \overline{B}_{\sqrt{m}}.$$

When $m = 2$, give a "proof by picture" of these set inequalities by drawing the three sets $\overline{B}_1$, $\overline{\text{Box}}_1$, and $\overline{B}_{\sqrt{2}}$.

2.9. The complex number system

Imagine a world in which we could not solve the equation $x^2 - 2 = 0$. This is a rational numbers only world. Such a world is a world where the length of the diagonal of a unit square would not make sense; a very poor world indeed! Imagine now a world in which every time we tried to solve a quadratic equation such as $x^2 + 1 = 0$, we get "stuck", and could not proceed further. This would incredibly slow down the progress of mathematics. The complex number system (introducing "imaginary numbers")[12] alleviates this potential stumbling block to mathematics and also to science . . . in fact, complex numbers are *necessary* to describe nature.[13]

[12] *The imaginary number is a fine and wonderful resource of the human spirit, almost an amphibian between being and not being. Gottfried Leibniz (1646–1716)* [**141**].

[13] *Furthermore, the use of complex numbers is in this case not a calculational trick of applied mathematics but comes close to being a necessity in the formulation of the laws of quantum mechanics ... It is difficult to avoid the impression that a miracle confronts us here. Nobel prize winner Eugene Wigner (1902–1995) responding to the "miraculous" appearance of complex numbers in the formulation of quantum mechanics* [**161**, p. 208], [**244**], [**245**].

2.9.1. Definition of complex numbers. The complex number system is actually very easy to define; it's really just $\mathbb{R}^2$! The **complex number system** $\mathbb{C}$ is the set $\mathbb{R}^2$ together with the following rules of arithmetic: If $z = (a, b)$ and $w = (c, d)$, then we already know how to add two such complex numbers:

$$z + w := (a, b) + (c, d) = (a + c, b + d);$$

the new ingredient is multiplication, which is defined by

$$\boxed{z \cdot w = (a, b) \cdot (c, d) := (ac - bd, ad + bc).}$$

In summary, $\mathbb{C}$ as a set is just $\mathbb{R}^2$, with the usual addition structure but with a special multiplication. Of course, we also define $-z = (-a, -b)$ and we write 0 for $(0, 0)$. Finally, if $z = (a, b) \neq 0$ (that is, $a \neq 0$ and $b \neq 0$), then we define

$$z^{-1} := \left(\frac{a}{a^2 + b^2} , \frac{-b}{a^2 + b^2} \right). \tag{2.34}$$

Theorem 2.42. *The complex numbers is a field with* $(0, 0)$ *(denoted henceforth by* 0*) and* $(1, 0)$ *(denoted henceforth by* 1*) being the additive and multiplicative identities, respectively.*

Proof. If $z, w, u \in \mathbb{C}$, then we need to show that addition satisfies

(A1) $z + w = w + z$; (commutative law)
(A2) $(z + w) + u = z + (w + u)$; (associative law)
(A3) $z + 0 = z = 0 + z$; (additive identity)
(A4) for each complex number z,

$$z + (-z) = 0 \quad \text{and} \quad (-z) + z = 0; \quad \text{(additive inverse)}$$

multiplication satisfies

(M1) $z \cdot w = w \cdot z$; (commutative law)
(M2) $(z \cdot w) \cdot u = z \cdot (w \cdot u)$; (associative law)
(M3) $1 \cdot z = z = z \cdot 1$; (multiplicative identity)
(M4) for $z \neq 0$, we have

$$z \cdot z^{-1} = 0 \quad \text{and} \quad z^{-1} \cdot z = 1; \quad \text{(multiplicative inverse)};$$

and finally, multiplication and addition are related by

(D) $z \cdot (w + u) = (z \cdot w) + (z \cdot u)$. (distributive law)

The proofs of all these properties are very easy and merely involve using the definition of addition and multiplication, so we leave all the proofs to the reader, except for (**M4**). Here, by definition of multiplication,

$$\begin{aligned} z \cdot z^{-1} &= (a, b) \cdot \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right) \\ &= \left(a \cdot \frac{a}{a^2 + b^2} \; - \; b \cdot \frac{-b}{a^2 + b^2}, \; a \cdot \frac{-b}{a^2 + b^2} \; + \; b \cdot \frac{a}{a^2 + b^2} \right) \\ &= \left(\frac{a^2}{a^2 + b^2} + \frac{b^2}{a^2 + b^2}, 0 \right) = (1, 0) = 1. \end{aligned}$$

Similarly, $z^{-1} \cdot z = 1$, and (**M4**) is proven. □

In particular, all the arithmetic properties of $\mathbb{R}$ hold for $\mathbb{C}$.

2.9.2. The number i. In high school, the complex numbers are introduced in a slightly different manner, which we now describe.

First, we consider $\mathbb{R}$ as a subset of $\mathbb{C}$ by the *identification* of the real number a with the ordered pair $(a, 0)$, in other words, for sake of notational convenience, we do not make a distinction between the complex number $(a, 0)$ and the real number a. Observe that by definition of addition and multiplication of complex numbers,

$$(a, 0) + (b, 0) = (a + b, 0)$$

and

$$(a, 0) \cdot (b, 0) = (a \cdot b - 0 \cdot 0, a \cdot 0 + 0 \cdot b) = (ab, 0),$$

which is to say, "$a+b = a+b$" and "$a \cdot b = a \cdot b$" under our identification. Thus, our identification of $\mathbb{R}$ preserves the arithmetic operations of $\mathbb{R}$. Moreover, we know how to multiply real numbers and elements of $\mathbb{R}^2$: $a(x, y) = (ax, ay)$. This also agrees with our complex number multiplication:

$$(a, 0) \cdot (x, y) = (a \cdot x - 0 \cdot y, a \cdot y + 0 \cdot x) = (ax, ay).$$

In summary, our identification of $\mathbb{R}$ as first components of ordered pairs in $\mathbb{C}$ does not harm any of the additive or multiplicative structures of $\mathbb{C}$.

The number i, notation introduced in 1777 by Euler [**171**], is by definition the complex number

$$\boxed{i := (0, 1).}$$

Then using the definition of multiplication of complex numbers, we have

$$i^2 = i \cdot i = (0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0) \quad \Longrightarrow \quad i^2 = -1,$$

where we used our identification of $(-1, 0)$ with -1. Thus, the complex number $i = (0, 1)$ is the "imaginary unit" that you learned about in high school; however, our definition of i avoids the mysterious square root of -1 you probably encountered.[14] Moreover, given any complex number $z = (a, b)$, by definition of addition, multiplication, i, and our identification of $\mathbb{R}$ as a subset of $\mathbb{C}$, we see that

$$\begin{aligned} a + b\,i = a + (b, 0) \cdot (0, 1) &= a + (b \cdot 0 - 0 \cdot 1, b \cdot 1 + 0 \cdot 0) \\ &= (a, 0) + (0, b) = (a, b) = z. \end{aligned}$$

Thus, $z = a + b\,i$, just as you were taught in high school! By commutativity, we also have $z = a + i\,b$. We call a the **real part** of z and b the **imaginary part** of z, and we denote them by $a = \operatorname{Re} z$ and $b = \operatorname{Im} z$, so that

$$\boxed{z = \operatorname{Re} z + i \operatorname{Im} z.}$$

From this point on, we shall typically use the notation $z = a + b\,i = a + i\,b$ instead of $z = (a, b)$ for complex numbers.

[14] *That this subject [imaginary numbers] has hitherto been surrounded by mysterious obscurity, is to be attributed largely to an ill adapted notation. If, for example, +1, -1, and the square root of -1 had been called direct, inverse and lateral units, instead of positive, negative and imaginary (or even impossible), such an obscurity would have been out of the question. Carl Friedrich Gauss (1777–1855).*

2.9.3. Absolute values and complex conjugates. We define the **absolute value** of a complex number $z = (a, b)$ (or **length**) as the usual length (or norm) of (a, b):

$$\boxed{|z| := |(a,b)| = \sqrt{a^2+b^2}.}$$

Thus, $|z|^2 = a^2 + b^2$. We define the **complex conjugate** of $z = (a, b)$ as the complex number $\overline{z} = (a, -b)$, that is, $\overline{z} = a - bi$. Note that if $|z| \neq 0$, then according to the definition (2.34) of z^{-1}, we have

$$z^{-1} = \frac{\overline{z}}{|z|^2},$$

so the inverse of a complex number can be expressed in terms of the complex conjugate and absolute value. In the next theorem we list other properties of the complex conjugate.

Theorem 2.43. *If z and w are complex numbers, then*

(1) $\overline{\overline{z}} = z$;
(2) $\overline{z+w} = \overline{z} + \overline{w}$ *and* $\overline{zw} = \overline{z} \cdot \overline{w}$;
(3) $z\,\overline{z} = |z|^2$;
(4) $z + \overline{z} = 2\operatorname{Re} z$ *and* $z - \overline{z} = 2i \operatorname{Im} z$.

Proof. The proofs of all these properties are very easy and merely involve using the definition of complex conjugation, so we leave all the proofs to the reader, except the last two. We have

$$\begin{aligned} z\,\overline{z} = (a+bi)\,(a-bi) &= a^2 + a(-bi) + (bi)a + (bi)(-bi) \\ &= a^2 - abi + abi - b^2 \cdot i^2 = a^2 - b^2 \cdot (-1) = a^2 + b^2 = |z|^2. \end{aligned}$$

To prove *(4)*, observe that $z + \overline{z} = a + bi + (a - bi) = 2a = 2\operatorname{Re} z$ and $z - \overline{z} = a + bi - (a - bi) = 2bi = 2i \operatorname{Im} z$. □

In the final theorem of this section we list various properties of absolute value.

Theorem 2.44. *For any complex numbers z, w, we have*

(i) $|z| \geq 0$ *and* $|z| = 0$ *if and only if* $z = 0$;
(ii) $|\overline{z}| = |z|$;
(iii) $|\operatorname{Re} z| \leq |z|$;
(iv) $|z\,w| = |z|\,|w|$;
(v) $|z + w| \leq |z| + |w|$. ***(triangle inequality)***

Proof. Properties *(i)* and *(v)* follow from the properties of the norm on $\mathbb{R}^2$. Properties *(ii)* and *(iii)* are straightforward to check. To prove *(iv)*, note that

$$|z\,w|^2 = \overline{zw}\,zw = \overline{z}\,\overline{w}\,zw = \overline{z}z\,\overline{w}w = |z|^2\,|w|^2.$$

Taking the square root of both sides shows that $|z\,w| = |z|\,|w|$. □

An induction argument shows that for any n complex numbers $z_1, \ldots, z_n$,

$$|z_1\, z_2 \cdots z_n| = |z_1|\,|z_2| \cdots |z_n|.$$

In particular, setting $z_1 = z_2 = \cdots = z_n = z$, we see that

$$|z^n| = |z|^n.$$

Exercises 2.9.

1. Show that $z \in \mathbb{C}$ is a real number if and only if $\overline{z} = z$.
2. If w is a complex root of a polynomial $p(z) = z^n + a_{n-1}\, z^{n-1} + \cdots + a_1 z + a_0$ with real coefficients (that is, $p(w) = 0$ and each a_k is real), prove that $\overline{w}$ is also a root.
3. If $z \in \mathbb{C}$, prove that there exists a nonnegative real number r and a complex number ω with $|\omega| = 1$ such that $z = r\,\omega$. If z is nonzero, show that r and ω are uniquely determined by z, that is, if $z = r'\,\omega'$ where $r' \geq 0$ and $|\omega'| = 1$, then $r' = r$ and $\omega' = \omega$. The decomposition $z = r\,\omega$ is called the **polar decomposition** of z. (In Section 4.7 we relate the polar decomposition to the trigonometric functions.)

2.10. Cardinality and "most" real numbers are transcendental

In Section 2.6 we have seen that in some sense (in dealing with roots, trig functions, logarithms — objects of practical interest) there are immeasurably more irrational numbers than there are rational numbers. This begs the question:[15] How much more? In this section we discuss Cantor's strange discovery that the rational numbers have in some sense the same number of elements as the natural numbers do! The rational numbers are thus said to be countable. It turns out that there are just as many irrational numbers are there are real numbers; the irrational and real numbers are said to be uncountable. We shall also discuss algebraic and transcendental numbers and discuss their countability properties.

2.10.1. Cardinality. Cardinality is simply a mathematical way to say that two sets have the same number of elements. Two sets A and B are said to have the same **cardinality**, if there is a bijection between these two sets. Of course, if $f : A \longrightarrow B$ is a bijection, then $g = f^{-1} : B \longrightarrow A$ is a bijection, so the notion of cardinality does not depend on "which way the bijection goes". We think of A and B as having the same number of elements since the bijection sets up a one-to-one correspondence between elements of the two sets. A set A is said to be **finite** if it is empty or if it has the same cardinality as a set of the form $\mathbb{N}_n := \{1, 2, \ldots, n\}$ for some natural number n, in which case we say that A has zero elements or n **elements**, respectively. If A is not finite, it is said to be **infinite**.[16] A set is called **countable** if it has the same cardinality as a finite set or the set of natural numbers. To distinguish between finite and infinite countable sets, we call a set **countably infinite** if it has the cardinality of the natural numbers. Finally, a set is **uncountable** if it is not countable, so the set is not finite and does not have the cardinality of $\mathbb{N}$. See Figure 2.11 for relationships between finite and countable sets. If $f : \mathbb{N} \longrightarrow A$ is a bijection, then A can be listed:

$$A = \{a_1, a_2, a_3, \ldots\},$$

where $a_n = f(n)$ for each $n = 1, 2, 3, \ldots$.

Example 2.29. The integers are countably infinite since the function $f : \mathbb{Z} \longrightarrow \mathbb{N}$ defined by

$$f(n) = \begin{cases} 2n & \text{if } n > 0 \\ 2|n| + 1 & \text{if } n \leq 0 \end{cases}$$

is a bijection of $\mathbb{Z}$ onto $\mathbb{N}$.

[15] *In mathematics the art of proposing a question must be held of higher value than solving it. (A thesis defended at Cantor's doctoral examination.) Georg Cantor (1845–1918).*

[16] *Even in the realm of things which do not claim actuality, and do not even claim possibility, there exist beyond dispute sets which are infinite. Bernard Bolzano (1781–1848).*

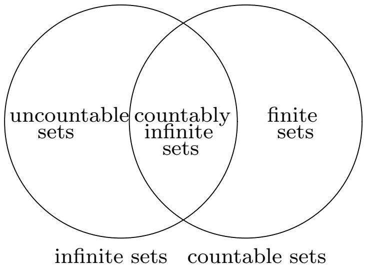


FIGURE 2.11. Infinite sets are uncountable or countably infinite and countable sets are countably infinite or finite. Infinite sets and countable sets intersect in the countably infinite sets.

If two sets A and B have the same cardinality, we sometimes write $\operatorname{card}(A) = \operatorname{card}(B)$. One can check that if $\operatorname{card}(A) = \operatorname{card}(B)$ and $\operatorname{card}(B) = \operatorname{card}(C)$, then $\operatorname{card}(A) = \operatorname{card}(C)$. Thus, cardinality satisfies a "transitive law".

It is "obvious" that a set cannot have both n elements and m elements where $n \neq m$, but this still needs proof! The proof is based on the "pigeonhole principle", which can be interpreted as saying that if $m > n$ and m pigeons are put into n holes, then at least two pigeons must be put into the same hole.

THEOREM 2.45 (**Pigeonhole principle**). *If $m > n$, then there does not exist an injection from $\mathbb{N}_m$ into $\mathbb{N}_n$.*

PROOF. We proceed by induction on n. Let $m > 1$ and $f : \mathbb{N}_m \longrightarrow \{1\}$ be any function. Then $f(m) = f(1) = 1$, so f is not an injection.

Assume that our theorem is true for n; we shall prove it true for $n + 1$. Let $m > n + 1$ and let $f : \mathbb{N}_m \longrightarrow \mathbb{N}_{n+1}$. We shall prove that f is not an injection. First of all, if the range of f is contained in $\mathbb{N}_n \subseteq \mathbb{N}_{n+1}$, then we can consider f as a function into $\mathbb{N}_n$, and hence by induction hypothesis, f is not an injection. So assume that the $f(a) = n + 1$ for some $a \in \mathbb{N}_m$. If there is another element of $\mathbb{N}_m$ whose image is $n + 1$, then f is not injection, so we may assume that a is the only element of $\mathbb{N}_m$ whose image is $n + 1$. Then $f(k) \in \mathbb{N}_n$ for $k \neq a$, so we can define a function $g : \mathbb{N}_{m-1} \longrightarrow \mathbb{N}_n$ by "skipping" $f(a) = n + 1$:

$$g(1) := f(1),\ g(2) := f(2), \ldots, g(a-1) := f(a-1),\ g(a) := f(a+1),$$
$$g(a+1) := f(a+2), \ldots, g(m-1) := f(m).$$

Since $m > n + 1$, we have $m - 1 > n$, so by induction hypothesis, g is not an injection. The definition of g shows that $f : \mathbb{N}_m \longrightarrow \mathbb{N}_{n+1}$ cannot be an injection either, which completes the proof of our theorem. □

We now prove that the number of elements of a finite set is unique. We also prove the "obvious" fact that an infinitely countable set is not finite.

THEOREM 2.46. *The number of elements of a finite set is unique and an infinitely countable set is not finite.*

PROOF. Suppose that $f : A \longrightarrow \mathbb{N}_n$ and $g : A \longrightarrow \mathbb{N}_m$ are bijections where $m > n$. Then

$$f \circ g^{-1} : \mathbb{N}_m \longrightarrow \mathbb{N}_n$$

is a bijection, and hence in particular an injection, an impossibility by the pigeonhole principle. This proves that the number of elements of a finite set is unique.

Now suppose that $f : A \longrightarrow \mathbb{N}_n$ and $g : A \longrightarrow \mathbb{N}$ are bijections. Then,

$$f \circ g^{-1} : \mathbb{N} \longrightarrow \mathbb{N}_n$$

is a bijection, so an injection, and so in particular, its restriction to $\mathbb{N}_{n+1} \subseteq \mathbb{N}$ is an injection. This again is impossible by the pigeonhole principle. □

2.10.2. Basic results on countability. The following is intuitively obvious.

Lemma 2.47. *A subset of a countable set is countable.*

Proof. Let A be a nonempty subset of a countable set B, where for definiteness we assume that B is countably infinite. (The finite case is left to the reader.) Let $f : \mathbb{N} \longrightarrow B$ be a bijection. Using the well-ordering principle, we can define

$$n_1 := \text{smallest element of } \{n \in \mathbb{N}\, ;\, f(n) \in A\}.$$

If $A \neq \{f(n_1)\}$, then via well-ordering, we can define

$$n_2 := \text{smallest element of } \{n \in \mathbb{N} \setminus \{n_1\}\, ;\, f(n) \in A\ \}.$$

Note that $n_1 < n_2$ (why?). If $A \neq \{f(n_1), f(n_2)\}$, then we can define

$$n_3 := \text{smallest element of } \{n \in \mathbb{N} \setminus \{n_1, n_2\}\, ;\, f(n) \in A\ \}.$$

Then $n_1 < n_2 < n_3$. We can continue this process by induction defining n_{k+1} as the smallest element in the set $\{n \in \mathbb{N} \setminus \{n_1, \ldots, n_k\}\, ;\, f(n) \in A\}$ as long as $A \neq \{f(n_1), \ldots, f(n_k)\}$.

There are two possibilities: the above process terminates or it continues indefinitely. If the process terminates, let n_m be the last natural number that can be defined in this process. Then $A = \{f(n_1), \ldots, f(n_m)\}$, which shows that $g : \mathbb{N}_m \longrightarrow A$ defined by $g(k) := f(n_k)$ is a bijection. If the above process can be continued indefinitely, we can produce an infinite sequence of natural numbers $n_1 < n_2 < n_3 < n_4 < \cdots$ using the above recursive procedure. Since $n_1 < n_2 < n_3 < \cdots$ is an increasing sequence of natural numbers, one can check (for instance, by induction) that $k \leq n_k$ for all k. We claim that the map $h : \mathbb{N} \longrightarrow A$ defined by $h(k) := f(n_k)$ is a bijection. It is certainly injective because f is. To see that h is surjective, let $a \in A$. Then, because f is surjective, there is an $\ell \in \mathbb{N}$ such that $f(\ell) = a$. Since $k \leq n_k$ for every k, by the Archimedean property, there is a k such that $\ell < n_{k+1}$. We claim that $\ell \in \{n_1, \ldots, n_k\}$. Indeed, if not, then $\ell \in \{n \in \mathbb{N} \setminus \{n_1, n_2, \ldots n_k\}\, ;\, f(n) \in A\}$, so by definition of n_{k+1},

$$n_{k+1} := \text{smallest element of } \{n \in \mathbb{N} \setminus \{n_1, n_2, \ldots n_k\}\, ;\, f(n) \in A\ \} \leq \ell,$$

contradicting that $\ell < n_{k+1}$. Hence, $\ell = n_j$ for some j, so $h(j) = f(n_j) = f(\ell) = a$. This proves that h is surjective and completes our proof. □

Theorem 2.48. *A finite product of countable sets is countable and a countable union of countable sets is countable.*

Proof. We only consider the product of two countably infinite sets (the other cases are left to the reader). The countability of the product of more than two countable sets can be handled by induction. If A and B are countably infinite, then $\operatorname{card}(A \times B) = \operatorname{card}(\mathbb{N} \times \mathbb{N})$, so it suffices to show that $\operatorname{card}(\mathbb{N} \times \mathbb{N}) = \operatorname{card}(\mathbb{N})$. Let $C \subseteq \mathbb{N}$ consist of all natural numbers of the form $2^n\, 3^m$ where $n, m \in \mathbb{N}$. Being an infinite subset of $\mathbb{N}$, it follows that C is countably infinite (that is, $\text{card}(C) = \text{card}(\mathbb{N})$). Consider the function $f : \mathbb{N} \times \mathbb{N} \longrightarrow C$ defined by

$$f(n,m) := 2^n \, 3^m.$$

By unique factorization, f is one-to-one, and so $\mathbb{N} \times \mathbb{N}$ has the same cardinality as the countably infinite set C (that is, $\text{card}(\mathbb{N} \times \mathbb{N}) = \text{card}(C) = \text{card}(\mathbb{N})$). Thus, $\mathbb{N} \times \mathbb{N}$ is countable. See Problem 1 for other proofs that $\mathbb{N} \times \mathbb{N}$ is countable.

Let $A = \bigcup_{n=1}^{\infty} A_n$ be a countable union of countable sets A_n. Since A_n is countable, we can list the (distinct) elements of A_n:

$$A_n = \{a_{n1}, a_{n2}, a_{n3}, \ldots\},$$

and each element a of A is of the form $a = a_{nm}$ for some pair (n,m), which may not be unique because a_{nm} and $a_{n'm'}$ could be the same for different (n,m) and (n',m'). To identify a unique such pair we require that n be the least such number with $a = a_{nm}$ for some m. This recipe defines a map $g : A \longrightarrow \mathbb{N} \times \mathbb{N}$ by

$$g(a) := (n,m),$$

which, as the reader can check, is one-to-one. So, A has the same cardinality as the subset $g(A)$ of the countable set $\mathbb{N} \times \mathbb{N}$. Since subsets of countable sets are countable, it follows that A has the same cardinality as a countable set, so A is countable. □

Example 2.30. (Cf. Example 2.29) As an easy application of this theorem, we observe that $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup (-\mathbb{N})$, and since each set on the right is countable, their union $\mathbb{Z}$ is also countable.

2.10.3. Real, rational, and irrational numbers.

We now prove that the rational numbers are countable.

Theorem 2.49. *The set of rational numbers is countably infinite.*

Proof. Let $A := \{(m,n) \in \mathbb{Z} \times \mathbb{N}\,;\, m \text{ and } n \text{ have no common factors}\}$. Since a product of countable sets is countable, $\mathbb{Z} \times \mathbb{N}$ is countable and since A is a subset of a countable set, A is countable. Moreover, A is infinite since, for example, all numbers of the form $(m,1)$ belong to A where $m \in \mathbb{Z}$. Define $f : A \longrightarrow \mathbb{Q}$ by

$$f(m,n) := \frac{m}{n}.$$

This function is a bijection, so $\text{card}(\mathbb{Q}) = \text{card}(A) = \text{card}(\mathbb{N})$. □

The following is Cantor's first proof that $\mathbb{R}$ is uncountable. (The following proof is close to, but not exactly, Cantor's original proof; see [**87**] for a nice exposition on his original proof.) His second proof is in Section 3.8.

Theorem 2.50 (**Cantor's first proof**). *Any interval of real numbers that is not empty or consisting of a single point is uncountable.*

Proof. Here we are omitting the empty interval and intervals of the form $[a,a] = \{a\}$. Suppose, for sake of contradiction, that there is such a countable interval: $I = \{c_1, c_2, \ldots\}$. Let $I_1 = [a_1, b_1] \subseteq I$, where $a_1 < b_1$, be an interval in I that does not contain c_1. (To see that such an interval exists, divide the interval I into three disjoint subintervals. At least one of the three subintervals does not contain c_1. Choose I_1 to be any closed interval in the subinterval that does not contain c_1.) Now let $I_2 = [a_2, b_2] \subseteq I_1$ be an interval that does not contain

c_2. By induction, we construct a sequence of nested closed and bounded intervals $I_n = [a_n, b_n]$ that does not contain c_n. By the nested intervals theorem, there is a point c in every I_n. By construction, I_n does not contain c_n, so c cannot equal any c_n, which contradicts that $\{c_1, c_2, \ldots\}$ is a list of all the real numbers in I. □

Example 2.31. With $I = \mathbb{R}$, we see that the set of all real numbers is uncountable. It follows that $\mathbb{R}^m$ is uncountable for any $m \in \mathbb{N}$; in particular, $\mathbb{C} = \mathbb{R}^2$ is uncountable.

Corollary 2.51. *The set of irrational numbers in any interval that is not empty or consisting of a single point is uncountable.*

Proof. If the irrationals in such an interval I were countable, then I would be the union of two countable sets, the irrationals in I and the rationals in I; however, we know that I is not countable so the irrationals in I cannot be countable. □

In particular, the set of all irrational numbers is uncountable.

2.10.4. Roots of polynomials. We already know that the real numbers are classified into two disjoint sets, the rational and irrational numbers. There is another important classification into algebraic and transcendental numbers. These numbers have to do with roots of polynomials, so we begin by discussing some elementary properties of polynomials. We remark that *everything* we say in this subsection and the next are valid for real polynomials (polynomials of a real variable with real coefficients), but it is convenient to work with complex polynomials.

Let $n \geq 1$ and

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_2 z^2 + a_1 z + a_0, \quad a_n \neq 0, \tag{2.35}$$

be an n-th degree polynomial with complex coefficients (that is, each $a_k \in \mathbb{C}$).

Lemma 2.52. *For any $z, a \in \mathbb{C}$, we can write*

$$p(z) - p(a) = (z - a)\, q(z),$$

where $q(z)$ is a polynomial of degree $n - 1$.

Proof. First of all, observe that given any polynomial $f(z)$ and a complex number b, the "shifted" function $f(z + b)$ is also a polynomial in z of the same degree as f; this can be easily proven using the formula (2.35) for a polynomial. In particular, $r(z) = p(z + a) - p(a)$ is a polynomial of degree n and hence can written in the form

$$r(z) = b_n z^n + b_{n-1} z^{n-1} + \cdots + b_2 z^2 + b_1 z + b_0,$$

where $b_n \neq 0$ (in fact, $b_n = a_n$ but this isn't needed). Notice that $r(0) = 0$, which implies that $b_0 = 0$, so we can write

$$r(z) = z\, s(z) \quad , \quad \text{where} \;\; s(z) = b_n z^{n-1} + b_{n-1} z^{n-2} + \cdots + b_2 z + b_1$$

is a polynomial of degree $n - 1$. Now replacing z with $z - a$, we obtain

$$p(z) - p(a) = r(z - a) = (z - a)\, q(z),$$

where $q(z) = s(z - a)$ is a polynomial of degree $n - 1$. □

Suppose that $a \in \mathbb{C}$ is a **root** of $p(z)$, which means $p(a) = 0$. Then according to our lemma, we can write $p(z) = (z - a)q(z)$ where q is a polynomial of degree $n - 1$ (here we drop the dependence on a in $q(z, a)$). If $q(a) = 0$, then again by our lemma, we can write $q(z) = (z - a)r(z)$ where $r(z)$ is a polynomial of degree $n - 2$. Thus, $p(z) = (z - a)^2 r(z)$. Continuing this process, which must stop by at least the n-th step (because the degree of a polynomial cannot be negative), we can write

$$p(z) = (z - a)^k s(z),$$

where $s(z)$ is a polynomial of degree $n - k$ and $s(a) \neq 0$. We say that a is a root of $p(z)$ of **multiplicity** k.

THEOREM 2.53. *Any n-th degree complex polynomial (see the expression* (2.35)*) has at most n complex roots counting multiplicity.*

PROOF. The proof is by induction. Certainly this theorem holds for polynomials of degree 1 (if $p(z) = a_1 z + a_0$ with $a_1 \neq 0$, then $p(z) = 0$ if and only if $z = -a_0/a_1$). Suppose that this theorem holds for polynomials of degree n. Let p be a polynomial of degree $n + 1$. If p has no roots, then this theorem holds for p, so suppose that p has a root, call it a. Then by our lemma we can write

$$p(z) = (z - a)q(z),$$

where q is a polynomial of degree n. By induction, we know that q has at most n roots counting multiplicity. The polynomial p has at most one more root (namely $z = a$) than q, so p has at most n roots counting multiplicity. □

By the fundamental theorem of algebra (Section 4.8) we'll see that any polynomial of degree n has exactly n (complex) roots counting multiplicities.

2.10.5. Uncountability of transcendental numbers. We already know that a rational number is a real number that can be written as a ratio of integers, and a number is irrational, by definition, if it is not rational. An important class of numbers that generalizes rational numbers is called the algebraic numbers. To motivate this generalization, let $r = a/b$, where $a, b \in \mathbb{Z}$ with $b \neq 0$, be a rational number. Then r is a root of the polynomial equation

$$bz - a = 0.$$

Therefore, any rational number is the root of a (**linear** or degree 1) polynomial with integer coefficients. In general, an **algebraic number** is a complex number that is a root of a polynomial with *integer* coefficients. A complex number is called **transcendental** if it is not algebraic. (These numbers are transcendental because as remarked by Euler, they "transcend" the power of algebra to solve for them.) Since complex numbers contain the real numbers, we can talk about real algebraic and transcendental numbers also. As demonstrated above, we already know that every rational number is algebraic. But there are many more algebraic numbers.

Example 2.32. $\sqrt{2}$ and $\sqrt[3]{5}$ are both algebraic, being roots of the polynomials

$$z^2 - 2 \quad \text{and} \quad z^3 - 5,$$

respectively. On the other hand, the numbers e and π are examples of transcendental numbers; for proofs see [**162**], [**163**], [**136**].

The numbers $\sqrt{2}$ and $\sqrt[3]{5}$ are irrational, so there are irrational numbers that are algebraic. Thus, the algebraic numbers include all rational numbers and infinitely many irrational numbers, namely those irrational numbers that are roots of polynomials with integer coefficients. Thus, it might seem as if the algebraic numbers are uncountable, while the transcendental numbers (those numbers that are not algebraic) are quite small in comparison. This is in fact not the case, as was discovered by Cantor.

THEOREM 2.54 (**Uncountability of transcendental numbers**). *The set of all algebraic numbers is countable and the set of all transcendental numbers is uncountable. The same statement holds for real algebraic and transcendental numbers.*

PROOF. We only consider the complex case. An algebraic number is by definition a complex number satisfying a polynomial equation with integer coefficients

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

Here, $n \geq 1$ (that is, this polynomial is *nonconstant*) in order for a solution to exist. We define the **index** of this polynomial as the number

$$n + |a_n| + |a_{n-1}| + |a_{n-2}| + \cdots + |a_2| + |a_1| + |a_0|.$$

Since $|a_n| \geq 1$, this index is at least 2 for any nonconstant polynomial with integer coefficients. Given any natural number k there are only a finite number of nonconstant polynomials with index k. For instance, there are only two nonconstant polynomials of index 2, the polynomials z and $-z$. There are eight nonconstant polynomials of index 3:

$$z^2, \quad z+1, \quad z-1, \quad 2z, \quad -z^2, \quad -z+1, \quad -z-1, \quad -2z,$$

and there are 22 polynomials of index 4:

$$z^3,\ 2z^2,\ z^2+z,\ ,z^2-z,\ z^2+1,\ z^2-1, 3z,\ 2z+1,\ 2z-1,\ z+2,\ z-2,$$

together with the negatives of these polynomials, and so forth. Since any polynomial of a given degree has finitely many roots (Theorem 2.53) and there are only a finite number of polynomials with a given index, the set A_k, consisting of all roots (algebraic numbers) of polynomials of index k, is a finite set. Since every polynomial with integer coefficients has an index, it follows that the set of all algebraic numbers is the union $\bigcup_{k=2}^{\infty} A_k$. Since a countable union of countable sets is countable, the set of algebraic numbers is countable!

We know that the complex numbers is the disjoint union of algebraic numbers and of transcendental numbers. Since the set of complex numbers is uncountable, the set of transcendental numbers must therefore be uncountable. □

EXERCISES 2.10.

1. Here are some countability proofs.
   (a) Prove that the set of prime numbers is countably infinite.
   (b) Let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Show that $\mathbb{N}_0$ is countably infinite. Define $f : \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0$ by $f(0,0) = 0$ and for $(m,n) \neq (0,0)$, define

   $$f(m,n) = 1 + 2 + 3 + \cdots + (m+n) + n = \frac{1}{2}(m+n)(m+n+1) + n.$$

   Can you see (do not prove) that this function counts $\mathbb{N}_0 \times \mathbb{N}_0$ as shown in Figure 2.12? Unfortunately, it is not so easy to show that f is a bijection.

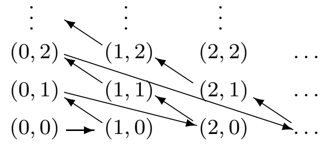


FIGURE 2.12. Visualization of the map $f : \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0$.

(c) Write $\mathbb{Q}$ as a countable union of countable sets, so giving another proof that the rational numbers are countable.

(d) Prove that $f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ defined by $f(m, n) = 2^{m-1}(2n-1)$ is a bijection; this give another proof that $\mathbb{N} \times \mathbb{N}$ is countable.

2. Here are some formulas for polynomials in terms of roots.

(a) If $c_1, \ldots, c_k$ are roots of a polynomial $p(z)$ of degree n (with each root repeated according to multiplicity), prove that $p(z) = (z - c_1)(z - c_2) \cdots (z - c_k)\, q(z)$, where $q(z)$ is a polynomial of degree $n - k$.

(b) If $k = n$, prove that $p(z) = a_n(z - c_1)(z - c_2) \cdots (z - c_n)$ where a_n is the coefficient of z^n in the formula (2.35) for $p(z)$.

3. Prove that if A is an infinite set, then A has a countably infinite subset.

4. Let X be any set and denote the set of all functions from X into $\{0, 1\}$ by $\mathbb{Z}_2^X$. Define a map from the power set of X into $\mathbb{Z}_2^X$ by

$$f : \mathscr{P}(X) \longrightarrow \mathbb{Z}_2^X, \qquad X \supseteq A \quad \longmapsto \quad f(A) := \chi_A,$$

where χ_A is the characteristic function of A. Prove that f is a bijection. Conclude that $\mathscr{P}(X)$ has the same cardinality as $\mathbb{Z}_2^X$.

5. Suppose that $\operatorname{card}(X) = n$. Prove that $\operatorname{card}(\mathscr{P}(X)) = 2^n$. Suggestion: There are many proofs you can come up with; here's one using the previous problem. Assuming that $X = \{0, 1, \ldots, n - 1\}$, which we may (why?), we just have to prove that $\operatorname{card}(\mathbb{Z}_2^X) = 2^n$. To prove this, define $F : \mathbb{Z}_2^X \longrightarrow \{0, 1, 2, \ldots, 2^n - 1\}$ as follows: If $f : X \longrightarrow \{0, 1\}$ is a function, then denoting $f(k)$ by a_k, define

$$F(f) := a_{n-1}\, 2^{n-1} + a_{n-2}\, 2^{n-2} + \cdots + a_1\, 2^1 + a_0.$$

Prove that F is a bijection (Section 2.5 will come in handy).

6. (**Cantor's theorem**) This theorem is simple to prove yet profound in nature.

(a) Prove that there can never be a surjection of a set A onto its power set $\mathscr{P}(A)$ (This is called Cantor's theorem). In particular, $\operatorname{card}(A) \neq \operatorname{card}(\mathscr{P}(A))$. Suggestion: Suppose not and let f be such a surjection. Consider the set

$$B = \{a \in A\,;\ a \notin f(a)\} \subseteq A.$$

Derive a contradiction from the assumption that f is surjective. Cantor's theorem shows that by taking power sets one can always get bigger and bigger sets.

(b) Prove that the set of all subsets of $\mathbb{N}$ is uncountable.

(c) From Cantor's theorem and Problem 4 prove that the set of all sequences of 0's and 1's is uncountable. Here, a **sequence** is just function from $\mathbb{N}$ into $\{0, 1\}$, which can also be thought of as a list $(a_1, a_2, a_3, a_4, \ldots)$ where each a_k is either 0 or 1.

7. (**Vredenduin's paradox** [**234**]) Here is another paradox related to the Russell's paradox. Assume that $A = \{\{a\}\,;\ a \text{ is a set}\}$ is a well-defined set. Let $B \subseteq A$

be the subset consisting of all sets of the form $\{a\}$ where $a \in \mathscr{P}(A)$. Define

$$g : \mathscr{P}(A) \longrightarrow B \quad \text{by} \quad g(V) := \{V\}.$$

Show that g is a bijection and then derive a contradiction to Cantor's theorem. This shows that A is not a set.

8. We define a **statement** as a finite string of symbols found on the common computer keyboard (we regard a space as a symbol). E.g. *Binghamton University is sooo great! Math is fun!* is a statement. Let's suppose there are 100 symbols on the common keyboard.
   (a) Let A be the set of all statements. What's the cardinality of A?
   (b) Is the set of all possible mathematical proofs countable? Why?

CHAPTER 3

Infinite sequences of real and complex numbers

Notable enough, however, are the controversies over the series 1 - 1 + 1 - 1 + 1 - ... whose sum was given by Leibniz as 1/2, although others disagree. ... Understanding of this question is to be sought in the word "sum"; this idea, if thus conceived — namely, the sum of a series is said to be that quantity to which it is brought closer as more terms of the series are taken — has relevance only for convergent series, and we should in general give up the idea of sum for divergent series.
Leonhard Euler (1707–1783).

Analysis is often described as the study of *infinite processes*, of which the study of sequences and series form the backbone. It is in dealing with the concept of "infinite" in infinite processes that makes analysis technically challenging. In fact, the subject of sequences is when real analysis becomes "really hard".

Let us consider the following infinite series that Euler mentioned:

$$s = 1 - 1 + 1 - 1 + 1 - 1 + 1 - 1 + \cdots .$$

Let's manipulate this infinite series without being too careful. First, we notice that

$$s = (1-1) + (1-1) + (1-1) + \cdots = 0 + 0 + 0 + \cdots = 0,$$

so $s = 0$. On the other hand,

$$s = 1 - (1-1) - (1-1) - (1-1) - \cdots = 1 - 0 - 0 - 0 - \cdots = 1,$$

so $s = 1$. Finally, we can get Leibniz's value of $1/2$ as follows:

$$\begin{aligned} 2s = 2 - 2 + 2 - 2 + \cdots &= 1 + 1 - 1 - 1 + 1 + 1 - 1 - 1 + \cdots \\ &= 1 + (1-1) - (1-1) + (1-1) - (1-1) + \cdots \\ &= 1 + 0 - 0 + 0 - 0 + \cdots = 1, \end{aligned}$$

so $s = 1/2$! This example shows us that we need to be careful in dealing with the infinite. In the pages that follow we "tame the infinite" with rigorous definitions.

Another highlight of this chapter is our study of the number e (Euler's number), which you have seen countless times in calculus and which pops up everywhere including economics (compound interest), population growth, radioactive decay, probability, etc. We shall prove two of the most famous formulas for this number:

$$\boxed{e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n .}$$

See [**55**], [**145**] for more on this incredible and versatile number. Another number we'll look at is the golden ratio $\Phi = \frac{1+\sqrt{5}}{2}$ which has strikingly pretty formulas

$$\boxed{\Phi = \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}} = 1 + \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}.}$$

In Section 3.1 we begin our study of infinite processes by learning about sequences and their limits, then in Section 3.2 we discuss the properties of sequences. Sections 3.3 and 3.4 are devoted to answering the question of when a given sequence converges; in these sections we'll also derive the above formulas for Φ. Next, in Section 3.5, we study infinite series, which is really a special case of the study of infinite sequences. The exponential function, called by many "the most important function in mathematics" [**192**, p. 1], is our subject of study in Section 3.7. This function is defined by

$$\boxed{\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad z \in \mathbb{C}.}$$

We shall derive a few of the exponential function's many properties including its relationship to Euler's number e. As a bonus prize, in Section 3.7 we'll also prove that e is irrational and we look at a useful (but little publicized) theorem called *Tannery's theorem*, which is a very handy result we'll used in subsequent sections. Finally, in Section 3.8 we see how real numbers can be represented as decimals (with respect to arbitrary bases) and we look at Cantor's famous "constructive" diagonal argument.

CHAPTER 3 OBJECTIVES: THE STUDENT WILL BE ABLE TO ...

- apply the rigorous ε-N definition of convergence for sequences and series.
- determine when a sequence is monotone, Cauchy, or has a convergent subsequence (Bolzano-Weierstrass), and when a series converges (absolutely).
- define the exponential function and the number e.
- explain Cantor's diagonal argument.

3.1. Convergence and ε-N arguments for limits of sequences

Undeniably, the most important concept in all of undergraduate analysis is the notion of convergence. Intuitively, a sequence $\{a_n\}$ in $\mathbb{R}^m$ converges to an element a in $\mathbb{R}^m$ indicates that a_n is "as close as we want" to a for n "sufficiently large". In this section we make the terms in quotes rigorous, which introduces the first bona fide technical definition in this book: the ε-N definition of limit.

3.1.1. Definition of convergence. A **sequence** in $\mathbb{R}^m$ can be thought of as a list

$$a_1,\ a_2,\ a_3,\ a_4,\ \ldots$$

of vectors, or points, a_n in $\mathbb{R}^m$. In the language of functions, a sequence is simply a function $f : \mathbb{N} \longrightarrow \mathbb{R}^m$, where we denote $f(n)$ by a_n. Usually a sequence is denoted by $\{a_n\}$ or by $\{a_n\}_{n=1}^{\infty}$. Of course, we are not restricted to $n \geq 1$ and we could start at any integer too, e.g. $\{a_n\}_{n=-5}^{\infty}$. For convenience, in most of our proofs we shall work with sequences starting at $n = 1$, although all the results we shall discuss work for sequences starting with any index.

Example 3.1. Some examples of real sequences (that is, sequences in $\mathbb{R}^1 = \mathbb{R}$) include[1]

$$3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415,\ \ldots$$

and

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \frac{1}{5},\ \frac{1}{6},\ \ldots, a_n = \frac{1}{n}, \ldots.$$

We are mostly interested in real or complex sequences. Here, by a complex sequence we simply mean a sequence in $\mathbb{R}^2$ where we are free to use the notation of i for $(0, 1)$ and the multiplicative structure.

Example 3.2. The following sequence is a complex sequence:

$$i,\ i^2 = -1,\ i^3 = -i,\ i^4 = 1,\ \ldots, a_n = i^n, \ldots$$

Although we shall focus on $\mathbb{R}$ and $\mathbb{C}$ sequences in this book, later on you might deal with topology and calculus in $\mathbb{R}^m$ (as in, for instance, [**136**]), so for your later psychological health we might as well get used to working with $\mathbb{R}^m$ instead of $\mathbb{R}^1$.

We now try to painstakingly motivate a precise definition of convergence (so please bear with me). Intuitively, a sequence $\{a_n\}$ in $\mathbb{R}^m$ converges to an element L in $\mathbb{R}^m$ indicates that a_n is "as close as we want" to L for n "sufficiently large". We now make the terms in quotes rigorous. First of all, what does "as close as we want" mean? We take it to mean that given any error, say $\varepsilon > 0$ (e.g. $\varepsilon = 0.01$), for n "sufficiently large" we can approximate L by a_n to within an error of ε. In other words, for n "sufficiently large" the difference between L and a_n is within ε:

$$|a_n - L| < \varepsilon.$$

Now what does for n "sufficiently large" mean? We define it to mean that there is a real number N such that for all $n > N$, a specified property holds (e.g. the above inequality); thus, for all $n > N$ we have $|a_n - L| < \varepsilon$, or using symbols,

$$n > N \implies |a_n - L| < \varepsilon. \tag{3.1}$$

In conclusion: For any given error $\varepsilon > 0$ there is an N such that (3.1) holds.[2]

We now summarize our findings as a precise definition. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$. We say that $\{a_n\}$ **converges** (or **tends**) to an element L in $\mathbb{R}^m$ if, for every $\varepsilon > 0$, there is an $N \in \mathbb{R}$ such that for all $n > N$, $|a_n - L| < \varepsilon$. Because this definition is so important, we display it: $\{a_n\}$ converges to L if,

$$\boxed{\text{for every } \varepsilon > 0, \text{ there is an } N \in \mathbb{R} \text{ such that} \quad n > N \implies |a_n - L| < \varepsilon.}$$

We call $\{a_n\}$ a **convergent sequence**, L the **limit** of $\{a_n\}$, and we usually denote the fact that $\{a_n\}$ converges to L in one of four ways:

$$a_n \to L, \qquad a_n \to L \text{ as } n \to \infty, \qquad \lim a_n = L, \qquad \lim_{n\to\infty} a_n = L.$$

If a sequence does not converge (to any element of $\mathbb{R}^m$), we say that it **diverges**.

We can also state the definition of convergence in terms of open balls. Observe

[1] We'll talk about decimal expansions of real numbers in Section 3.8 and π in Chapter 4.

[2] *One magnitude is said to be the limit of another magnitude when the second may approach the first within any given magnitude, however small, though the second may never exceed the magnitude it approaches. Jean Le Rond d'Alembert (1717–1783). The article on Limite in the Encyclopdie 1754.*

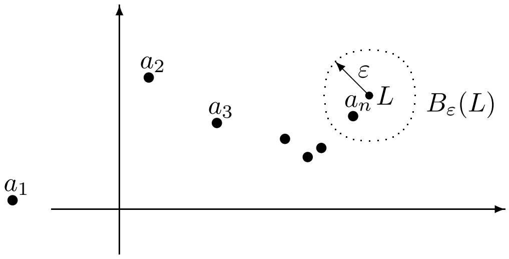


FIGURE 3.1. $a_n \to L$ if and only if given any ε-ball around L, the a_n's are "eventually" inside of the ε-ball.

that $|a_n - L| < \varepsilon$ is just saying that $a_n \in B_\varepsilon(L)$, the open ball of radius ε centered at L (see formula (2.32) in Section 2.8). Therefore, $a_n \to L$ in $\mathbb{R}^m$ if,

$$\boxed{\text{for every } \varepsilon > 0, \text{ there is an } N \in \mathbb{R} \text{ such that } \quad n > N \implies a_n \in B_\varepsilon(L).}$$

See Figure 3.1. We'll not emphasize this interpretation of limit but we state it because the open ball idea will occur in other classes, in particular, when studying metric spaces in topology.

3.1.2. Standard examples of ε-N arguments.

We now give some standard examples of using our "ε-N" definition of limit.

Example 3.3. We shall prove that the sequence $\{1/2, 1/3, 1/4, \ldots\}$ converges to zero:

$$\lim \frac{1}{n+1} = 0.$$

In general, any sequence $\{a_n\}$ that converges to zero is called a **null sequence**. Thus, we claim that $\{1/(n+1)\}$ is a null sequence. Let $\varepsilon > 0$ be any given positive real number. We want to prove there exists a real number N such that

$$n > N \implies \left| \frac{1}{n+1} - 0 \right| = \frac{1}{n+1} < \varepsilon. \tag{3.2}$$

To find such a number N, we can proceed in many ways. Here are two ways.

(I) For our first way, we observe that

$$\frac{1}{n+1} < \varepsilon \iff \frac{1}{\varepsilon} < n+1 \iff \frac{1}{\varepsilon} - 1 < n. \tag{3.3}$$

For this reason, let us choose N to be the real number $N = 1/\varepsilon - 1$. Let $n > N$, that is, $N < n$ or using the definition of N, $1/\varepsilon - 1 < n$. Then by (3.3), we have $1/(n+1) < \varepsilon$. In summary, for $n > N$, we have proved that $1/(n+1) < \varepsilon$. This proves (3.2). Thus, by definition of convergence, $1/(n+1) \to 0$.

(II) Another technique is to try and simplify the right-hand side of (3.2). Since $n < n+1$, we have $1/(n+1) < 1/n$. Therefore,

$$\text{if } \frac{1}{n} < \varepsilon, \text{ then because } \frac{1}{n+1} < \frac{1}{n}, \text{ we have } \frac{1}{n+1} < \varepsilon \text{ also.} \tag{3.4}$$

Now we can make $1/n < \varepsilon$ easily since

$$\frac{1}{n} < \varepsilon \iff \frac{1}{\varepsilon} < n.$$

With this scratch work done, let us now choose $N = 1/\varepsilon$. Let $n > N$, that is, $N < n$ or using the definition of N, $1/\varepsilon < n$. Then we certainly have $1/n < \varepsilon$, and hence by (3.4), we know that $1/(n+1) < \varepsilon$ too. In summary, for $n > N$, we have proved that $1/(n+1) < \varepsilon$. This proves (3.2).

Note that in (**I**) and (**II**), we found different N's (namely $N = 1/\varepsilon - 1$ in (**I**) and $N = 1/\varepsilon$ in (**II**)), but this doesn't matter because to prove (3.2) we just need to show such an N *exists*; it doesn't have to be unique and in general, many different N's will work. We remark that a similar argument shows that the sequence $\{1/n\}$ is also a null sequence: $\lim \frac{1}{n} = 0$.

Example 3.4. Here's a harder example. Let's prove that

$$\lim \frac{2n^2 - n}{n^2 - 9} = 2.$$

For the sequence $a_n = (2n^2 - n)/(n^2 - 9)$, we take the indices to be $n = 4, 5, 6, \ldots$ (since for $n = 3$ the quotient is undefined). Let $\varepsilon > 0$ be given. We want to prove there exists a real number N such that the following statement holds:

$$n > N \quad \Longrightarrow \quad \left| \frac{2n^2 - n}{n^2 - 9} - 2 \right| < \varepsilon.$$

One technique to prove this is to try and "massage" (simplify) the absolute value on the right as much as we can. For instance, we first can combine fractions:

$$\left| \frac{2n^2 - n}{n^2 - 9} - 2 \right| = \left| \frac{2n^2 - n}{n^2 - 9} - \frac{2n^2 - 18}{n^2 - 9} \right| = \left| \frac{18 - n}{n^2 - 9} \right|. \tag{3.5}$$

Second, just so that we don't have to worry about absolute values, we can get rid of them by using the triangle inequality: for $n = 4, 5, \ldots$, we have

$$\left| \frac{18 - n}{n^2 - 9} \right| \le \frac{18 + n}{n^2 - 9}. \tag{3.6}$$

Third, just for topping on the cake, let us make the top of the right-hand fraction a little simpler by observing that $18 \le 18n$, so we conclude that

$$\frac{18 + n}{n^2 - 9} \le \frac{18n + n}{n^2 - 9} = \frac{19n}{n^2 - 9}. \tag{3.7}$$

In conclusion, we have "massaged" our expression to the following inequality:

$$\left| \frac{2n^2 - n}{n^2 - 9} - 2 \right| \le \frac{19n}{n^2 - 9}.$$

So we just need to prove there is an N such that

$$n > N \quad \Longrightarrow \quad \frac{19n}{n^2 - 9} < \varepsilon; \tag{3.8}$$

this will automatically imply that for $n > N$, we have $\left| \frac{2n^2 - n}{n^2 - 9} - 2 \right| < \varepsilon$, which was what we originally wanted. There are many "tricks" to find an N satisfying (3.8). Here are three slightly different ways.

(I) For our first method, we use a technique sometimes found in elementary calculus: We divide top and bottom by n^2:

$$\frac{19n}{n^2 - 9} = \frac{19/n}{1 - 9/n^2}.$$

To show that this can be made less than ε, we need to show that the denominator can't get too small (otherwise the fraction $(19/n)/(1-9/n^2)$ might get large). To this end, observe that $\frac{9}{n^2} \leq \frac{9}{4^2}$ for $n \geq 4$, so

$$\text{for } n \geq 4, \quad 1 - \frac{9}{n^2} \geq 1 - \frac{9}{4^2} = 1 - \frac{9}{16} = \frac{7}{16} > \frac{1}{3}.$$

Hence, for $n \geq 4$, we have $\frac{1}{3} < 1 - \frac{9}{n^2}$, which is to say, $\frac{1}{1-9/n^2} < 3$. Thus,

$$\text{for } n \geq 4, \quad \frac{19n}{n^2-9} = \frac{19/n}{1-9/n^2} < (19/n) \cdot 3 = \frac{57}{n}. \tag{3.9}$$

Therefore, we can satisfy (3.8) by making $57/n < \varepsilon$ instead. Now,

$$\frac{57}{n} < \varepsilon \iff \frac{57}{\varepsilon} < n. \tag{3.10}$$

Because of (3.9) and (3.10), let us pick N to be the larger of 3 and $57/\varepsilon$. We'll prove that this N works for (3.8). Let $n > N$, which implies that $n > 3$ and $n > 57/\varepsilon$. In particular, $n \geq 4$ and $\varepsilon > 57/n$. Therefore,

$$\frac{19n}{n^2-9} \overset{\text{by (3.9)}}{<} \frac{57}{n} \overset{\text{by (3.10)}}{<} \varepsilon.$$

This proves (3.8).

(II) For our second method, observe that $n^2 - 9 \geq n^2 - 9n$ since $9 \leq 9n$. Hence,

$$\text{for } n > 9, \quad \frac{1}{n^2-9} \leq \frac{1}{n^2-9n} = \frac{1}{n(n-9)},$$

where we chose $n > 9$ so that $n(n-9)$ is positive. Thus,

$$\text{for } n > 9, \quad \frac{19n}{n^2-9} \leq \frac{19n}{n(n-9)} = \frac{19}{n-9}. \tag{3.11}$$

So, we can satisfy (3.8) by making $19/(n-9) < \varepsilon$ instead. Now, for $n > 9$,

$$\frac{19}{n-9} < \varepsilon \iff \frac{19}{\varepsilon} < n - 9 \iff 9 + \frac{19}{\varepsilon} < n. \tag{3.12}$$

Because of (3.11) and (3.12), let us pick $N = 9 + \frac{19}{\varepsilon}$. We'll prove that this N works for (3.8). Let $n > N$, which implies, in particular, that $n > 9$. Therefore,

$$\frac{19n}{n^2-9} \overset{\text{by (3.11)}}{<} \frac{19}{n-9} \overset{\text{by (3.12)}}{<} \varepsilon.$$

This proves (3.8).

(III) For our last (and my favorite of the three) method, we factor the bottom:

$$\frac{19n}{n^2-9} = \frac{19n}{(n+3)(n-3)} = \frac{19n}{n+3} \cdot \frac{1}{n-3} < 19 \cdot \frac{1}{n-3}, \tag{3.13}$$

where used the fact that $\frac{19n}{n+3} < \frac{19n}{n} = 19$. Now (solving for n),

$$\frac{19}{n-3} < \varepsilon \iff 3 + \frac{19}{\varepsilon} < n. \tag{3.14}$$

For this reason, let us pick $N = 3 + 19/\varepsilon$. We'll show that this N satisfies (3.8). Indeed, for $n > N$ (that is, $3 + 19/\varepsilon < n$), we have

$$\frac{19n}{n^2-9} \overset{\text{by (3.13)}}{<} \frac{19}{n-3} \overset{\text{by (3.14)}}{<} \varepsilon.$$

3.1.3. Sophisticated examples of ε-N arguments.

We now give some very famous classical examples of ε-N arguments.

Example 3.5. Let a be any complex number with $|a| < 1$ and consider the sequence $a, a^2, a^3, \ldots$ (so that $a_n = a^n$ for each n). We shall prove that $\{a_n\}$ is a null sequence, that is,

$$\boxed{\lim a^n = 0, \qquad |a| < 1.}$$

Let $\varepsilon > 0$ be any given positive real number. We need to prove that there is a real number N such that the following statement holds:

$$n > N \implies |a^n - 0| = |a|^n < \varepsilon. \tag{3.15}$$

If $a = 0$, then any N would do, so we might as well assume that $a \neq 0$. In this case, since the real number $|a|$ is less than 1, we can write $|a| = \frac{1}{1+b}$, where $b > 0$; in fact, we can simply take $b = -1 + 1/|a|$. (Since $|a| < 1$, we have $1/|a| > 1$, so $b > 0$.) Therefore,

$$|a|^n = \frac{1}{(1+b)^n}, \quad \text{where } b > 0.$$

By Bernoulli's inequality (Theorem 2.7),

$$(1+b)^n \geq 1 + nb \geq nb \implies \frac{1}{(1+b)^n} \leq \frac{1}{nb}.$$

Hence,

$$|a|^n \leq \frac{1}{nb}. \tag{3.16}$$

Thus, we can satisfy (3.15) by making $1/(nb) < \varepsilon$ instead. Now,

$$\frac{1}{nb} < \varepsilon \iff \frac{1}{b\varepsilon} < n. \tag{3.17}$$

For this reason, let us pick $N = 1/(b\varepsilon)$. Let $n > N$ (that is, $1/(b\varepsilon) < n$). Then,

$$|a|^n \overset{\text{by (3.16)}}{\leq} \frac{1}{nb} \overset{\text{by (3.17)}}{<} \varepsilon.$$

This proves (3.15) and thus, by definition of convergence, $a^n \to 0$.

Example 3.6. For our next example, let $a > 0$ be any positive real number and consider the sequence $a, a^{1/2}, a^{1/3}, \ldots$ (so that $a_n = a^{1/n}$ for each n). We shall prove that $a_n \to 1$, that is,

$$\boxed{\lim a^{1/n} = 1, \qquad a > 0.}$$

If $a = 1$, then the sequence $a^{1/n}$ is just the constant sequence $1, 1, 1, 1, \ldots$, which certainly converges to 1 (can you prove this?). Suppose that $a > 1$; we shall consider the case $0 < a < 1$ afterwards. Let $\varepsilon > 0$ be any given positive real number. We need to prove that there is a real number N such that

$$n > N \implies \left|a^{1/n} - 1\right| < \varepsilon. \tag{3.18}$$

By our familiar root rules (Theorem 2.32), we know that $a^{1/n} > 1^{1/n} = 1$ and therefore $b_n := a^{1/n} - 1 > 0$. By Bernoulli's inequality (Theorem 2.7), we have

$$a = \left(a^{1/n}\right)^n = (1 + b_n)^n \geq 1 + nb_n \geq nb_n \implies b_n \leq \frac{a}{n}.$$

Hence,

$$\left|a^{1/n} - 1\right| = |b_n| \le \frac{a}{n}. \tag{3.19}$$

Thus, we can satisfy (3.18) by making $a/n < \varepsilon$ instead. Now,

$$\frac{a}{n} < \varepsilon \quad \Longleftrightarrow \quad \frac{a}{\varepsilon} < n. \tag{3.20}$$

For this reason, let us pick $N = a/\varepsilon$. Let $n > N$ (that is, $a/\varepsilon < n$). Then,

$$\left|a^{1/n} - 1\right| \overset{\text{by (3.19)}}{\le} \frac{a}{n} \overset{\text{by (3.20)}}{<} \varepsilon.$$

So, by definition of convergence, $a^{1/n} \to 1$. Now consider the case when $0 < a < 1$. Let $\varepsilon > 0$ be any given positive real number. We need to prove that there is a real number N such that

$$n > N \quad \Longrightarrow \quad \left|a^{1/n} - 1\right| < \varepsilon.$$

Since $0 < a < 1$, we have $1/a > 1$, so by our argument for real numbers greater than one we know that $1/a^{1/n} = (1/a)^{1/n} \to 1$. Thus, there is a real number N such that

$$n > N \quad \Longrightarrow \quad \left|\frac{1}{a^{1/n}} - 1\right| < \varepsilon.$$

Multiplying both sides of the right-hand inequality by the positive real number $a^{1/n}$, we get $n > N \Longrightarrow \left|1 - a^{1/n}\right| < a^{1/n}\varepsilon$. Since $0 < a < 1$, by our root rules, $a^{1/n} < 1^{1/n} = 1$, so $a^{1/n}\varepsilon < 1 \cdot \varepsilon = \varepsilon$. Hence,

$$n > N \quad \Longrightarrow \quad \left|a^{1/n} - 1\right| < \varepsilon,$$

which shows that $a^{1/n} \to 0$ as we wished to show.

Example 3.7. We come to our last example, which may seem surprising at first. Consider the sequence $a_n = n^{1/n}$. We already know that if $a > 0$ is any fixed real number, then $a^{1/n} \to 1$. In our present case $a_n = n^{1/n}$, so the "a" is increasing with n and it is not at all obvious what $n^{1/n}$ converges to, if anything! However, we shall prove that

$$\boxed{\lim n^{1/n} = 1.}$$

For $n > 1$ by our root rules we know that $n^{1/n} > 1^{1/n} = 1$, so for $n > 1$, we conclude that $b_n := n^{1/n} - 1 > 0$. By the binomial theorem (Theorem 2.8), we have

$$n = (n^{1/n})^n = (1 + b_n)^n = 1 + \binom{n}{1} b_n + \binom{n}{2} b_n^2 + \cdots + \binom{n}{n} b_n^n.$$

Since $b_n > 0$, all the terms on the right-hand side are positive, so dropping all the terms except the third term on the right, we see that for $n > 1$,

$$n > \binom{n}{2} b_n^2 = \frac{n!}{2!\,(n-2)!}\, b_n^2 = \frac{n(n-1)}{2}\, b_n^2.$$

Cancelling off the n's from both sides, we obtain for $n > 1$,

$$b_n^2 < \frac{2}{n-1} \quad \Longrightarrow \quad b_n < \frac{\sqrt{2}}{\sqrt{n-1}}.$$

Hence, for $n > 1$,

$$\left|n^{1/n} - 1\right| = |b_n| < \frac{\sqrt{2}}{\sqrt{n-1}}. \tag{3.21}$$

Let $\varepsilon > 0$ be given. Then

$$\frac{\sqrt{2}}{\sqrt{n-1}} < \varepsilon \quad \Longleftrightarrow \quad 1 + \frac{2}{\varepsilon^2} < n. \tag{3.22}$$

For this reason, let us pick $N = 1 + 2/\varepsilon^2$. Let $n > N$ (that is, $1 + 2/\varepsilon^2 < n$). Then,

$$\left|n^{1/n} - 1\right| \overset{\text{by (3.21)}}{\leq} \frac{\sqrt{2}}{\sqrt{n-1}} \overset{\text{by (3.22)}}{<} \varepsilon.$$

Thus, by definition of convergence, $n^{1/n} \to 1$.

Exercises 3.1.

1. Using the ε-N definition of limit, prove that

$$(a)\ \lim \frac{(-1)^n}{n} = 0, \quad (b)\ \lim \left(2 + \frac{3}{n}\right) = 2, \quad (c)\ \lim \frac{n}{n-1} = 1, \quad (d)\ \lim \frac{(-1)^n}{\sqrt{n} - 1} = 0.$$

2. Using the ε-N definition of limit, prove that

$$(a)\ \lim \frac{5n^2 + 2}{n^3 - 3n + 1} = 0, \quad (b)\ \lim \frac{n^2 - \sqrt{n}}{3n^2 - 2} = \frac{1}{3}, \quad (c)\ \lim \left[\sqrt{n^2 + n} - n\right] = \frac{1}{2}.$$

3. Here is another method to prove that $a^{1/n} \to 1$.
   (i) Recall that for any real number b, $b^n - 1 = (b-1)(b^{n-1} + b^{n-2} + \cdots + b + 1)$. Using this formula, prove that if $a \geq 1$, then $a - 1 \geq n(a^{1/n} - 1)$. Suggestion: Let $b = a^{1/n}$.
   (ii) Now prove that for any $a > 0$, $a^{1/n} \to 1$. (Do the case $a \geq 1$ first, then consider the case $0 < a < 1$.)
   (iii) In fact, we shall prove that if $\{r_n\}$ is a sequence of rational numbers converging to zero, then for any $a > 0$, $a^{r_n} \to 1$. First, using (i), prove that if $a \geq 1$ and $0 < r < 1$ is rational, then $a^r - 1 \leq ra(a-1)$. Second, prove that if $a \geq 1$ and $0 < r < 1$ is rational, then $|a^{-r} - 1| \leq a^r - 1$. Using these facts, prove that for any $a > 0$, $a^{r_n} \to 1$. (As before, do the case $a \geq 1$ first, then $0 < a < 1$.)

4. Let a be a complex number with $|a| < 1$. We already know that $a^n \to 0$. In this problem we prove the, somewhat surprising, fact that $na^n \to 0$. Although n grows very large, this limit shows that a^n must go to zero faster than n grows.
   (i) As in Example 3.5 we can write $|a| = 1/(1+b)$ where $b > 0$. Using the binomial theorem, show that for $n > 1$,
   $$|a|^n < \frac{1}{\binom{n}{2} b^2}.$$
   (ii) Show that
   $$n|a|^n < \frac{2}{(n-1)\, b^2}.$$
   (iii) Now prove that $na^n \to 0$.

5. Here's an even more surprising fact. Let a be a complex number with $|a| < 1$. Prove that given any natural number $k > 0$, we have $n^k a^n \to 0$. Suggestion: Let $\alpha := |a|^{1/k} < 1$ and use the fact that $n\alpha^n \to 0$ by the previous problem.

6. If $\{a_n\}$ is a sequence of nonnegative real numbers and $a_n \to L$, prove
   (i) $L \geq 0$.
   (ii) $\sqrt{a_n} \to \sqrt{L}$. (You need to consider two cases, $L = 0$ and $L > 0$.)

7. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$ and let $L \in \mathbb{R}^m$. Form the negation of the definition that $a_n \to L$, thus giving a statement that $a_n \not\to L$ (the sequence $\{a_n\}$ does not tend to L). Using your negation, prove that the sequence $\{(-1)^n\}$ diverges, that is, does not converge to any real number. In the next section we shall find an easy way to verify that a sequence diverges using the notion of subsequences.
8. (**Infinite products** — see Chapter 7 for more on this amazing topic!) In this problem we investigate the **infinite product**

$$\frac{2^2}{1\cdot 3}\cdot\frac{3^2}{2\cdot 4}\cdot\frac{4^2}{3\cdot 5}\cdot\frac{5^2}{4\cdot 6}\cdot\frac{6^2}{5\cdot 7}\cdot\frac{7^2}{6\cdot 8}\cdots \tag{3.23}$$

We interpret this "infinite product" as the limit of the "partial products"

$$a_1 = \frac{2^2}{1\cdot 3},\quad a_2 = \frac{2^2}{1\cdot 3}\cdot\frac{3^2}{2\cdot 4},\quad a_3 = \frac{2^2}{1\cdot 3}\cdot\frac{3^2}{2\cdot 4}\cdot\frac{4^2}{3\cdot 5},\ldots.$$

In other words, for each $n \in \mathbb{N}$, we define $a_n := \frac{2^2}{1\cdot 3}\cdot\frac{3^2}{2\cdot 4}\cdots\frac{(n+1)^2}{n\cdot(n+2)}$. We prove that the sequence $\{a_n\}$ converges as follows.

(i) Prove that $a_n = \frac{2(n+1)}{n+2}$.

(ii) Now prove that $a_n \to 2$. We sometimes write the infinite product (3.23) using $\prod$ notation and we express the limit $\lim a_n = 2$ as

$$\frac{2^2}{1\cdot 3}\cdot\frac{3^2}{2\cdot 4}\cdot\frac{4^2}{3\cdot 5}\cdot\frac{5^2}{4\cdot 6}\cdots = 2 \quad\text{or}\quad \prod_{n=1}^{\infty}\frac{(n+1)^2}{n(n+2)} = 2.$$

3.2. A potpourri of limit properties for sequences

Now that we have a working knowledge of the ε-N definition of limit, we move onto studying the properties of limits that will be used throughout the rest of the book. In particular, we learn the "algebra of limits," which allows us to combine convergent sequences to form other convergent sequences. Finally, we discuss the notion of properly divergent sequences.

3.2.1. Some limit theorems.

We begin by proving that limits are unique, that is, a convergent sequence cannot have two different limits. Before doing so, we first prove a lemma.

Lemma 3.1 (**The ε-principle**). *If $x \in \mathbb{R}$ and for any $\varepsilon > 0$, we have $x \le \varepsilon$, then $x \le 0$. In particular, if $a \in \mathbb{R}^m$ and for any $\varepsilon > 0$, we have $|a| < \varepsilon$, then $a = 0$.*

Proof. By way of contradiction, assume that $x > 0$. Then choosing $\varepsilon = x/2 > 0$, by assumption we have $x < \varepsilon = x/2$. Subtracting $x/2$ from both sides, we conclude that $x/2 < 0$, which implies that $x < 0$, a contradiction.

The second assertion of the theorem follows by applying the first assertion to $x = |a|$. In this case, $|a| \le 0$, which implies that $|a| = 0$ and therefore $a = 0$. □

The proof of the following theorem uses the renowned "$\varepsilon/2$-trick."

Theorem 3.2 (**Uniqueness of limits**). *Sequences can have at most one limit.*

Proof. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$ and suppose that $a_n \to L$ and $a_n \to L'$. We shall prove that $L = L'$. To see this let $\varepsilon > 0$. Since $a_n \to L$, with $\varepsilon/2$ replacing ε in the definition of limit, there is an N such that $|a_n - L| < \varepsilon/2$ for all $n > N$.

Since $a_n \to L'$, with $\varepsilon/2$ replacing ε in the definition of limit, there is an N' such that $|a_n - L'| < \varepsilon/2$ for all $n > N'$. By the triangle inequality,

$$|L - L'| = |(L - a_n) + (a_n - L')| \leq |L - a_n| + |a_n - L'|.$$

In particular, for n greater than the larger of N and N', we see that

$$|L - L'| \leq |L - a_n| + |a_n - L'| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon;$$

that is, $|L - L'| < \varepsilon$. Now $\varepsilon > 0$ was completely arbitrary, so by the ε-principle, we must have $L - L' = 0$, or $L = L'$, which is what we intended to show. □

It is important that the convergence or divergence of a sequence depends only on the "tail" of the sequence, that is, on the terms of the sequence for large n. This fact is more-or-less obvious

Example 3.8. The sequence

$$-100,\ 100,\ 50,\ 1000,\ \frac{1}{2},\ \frac{1}{4},\ \frac{1}{8},\ \frac{1}{16},\ \frac{1}{32},\ \frac{1}{64},\ \frac{1}{128},\ \cdots$$

converges to zero, and the first terms of this sequence don't effect this fact.

Given a sequence $\{a_n\}$ in $\mathbb{R}^m$ and a nonnegative integer $k = 0, 1, 2, \ldots$, we call the sequence $\{a_{k+1}, a_{k+2}, a_{k+3}, a_{k+4}, \ldots\}$ a k-**tail** (of the sequence $\{a_n\}$). We'll leave the following proof to the reader.

THEOREM 3.3 (**Tails theorem for sequences**). *A sequence converges if and only if every tail converges, if and only if some tail converges.*

We now show that convergence in $\mathbb{R}^m$ can be reduced to convergence in $\mathbb{R}$ — this is why $\mathbb{R}$ sequences are so important. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$. Since $a_n \in \mathbb{R}^m$, we can express a_n in terms of its m-tuple of components:

$$a_n = (a_{1n}, a_{2n}, \ldots, a_{mn}).$$

Notice that each coordinate, say the k-th one a_{kn}, is a real number and so, $\{a_{kn}\} = \{a_{k1}, a_{k2}, a_{k3}, \ldots\}$ is a sequence in $\mathbb{R}$. Given $L = (L_1, L_2, \ldots, L_m) \in \mathbb{R}^m$, in the following theorem we prove that if $a_n \to L$, then for each $k = 1, \ldots, m$, $a_{kn} \to L_k$ as $n \to \infty$ as well. Conversely, we shall prove that if for each $k = 1, \ldots, m$, $a_{kn} \to L_k$ as $n \to \infty$, then $a_n \to L$ as well.

THEOREM 3.4 (**Component theorem**). *A sequence in $\mathbb{R}^m$ converges to $L \in \mathbb{R}^m$ if and only if each component sequence converges in $\mathbb{R}$ to the corresponding component of L.*

PROOF. Suppose first that $a_n \to L$. Fixing k, we shall prove that $a_{kn} \to L_k$. Let $\varepsilon > 0$. Since $a_n \to L$, there is an N such that for all $n > N$, $|a_n - L| < \varepsilon$. Hence, by definition of the norm on $\mathbb{R}^m$, for all $n > N$,

$$(a_{kn} - L_k)^2 \leq (a_{1n} - L_1)^2 + (a_{2n} - L_2)^2 + \cdots + (a_{mn} - L_m)^2 = |a_n - L|^2 < \varepsilon^2.$$

Taking square roots of both sides shows that for all $n > N$, $|a_{kn} - L_k| < \varepsilon$, which shows that $a_{kn} \to L_k$.

Suppose now that for each $k = 1, \ldots, m$, $a_{kn} \to L_k$. Let $\varepsilon > 0$. Since $a_{kn} \to L_k$ there is an N_k such that for all $n > N_k$, $|a_{kn} - L_k| < \varepsilon/\sqrt{m}$. Let N be the largest of the numbers $N_1, N_2, \ldots, N_m$. Then for $n > N$, we have

$$\begin{aligned}|a_n - L|^2 &= (a_{1n} - L_1)^2 + (a_{2n} - L_2)^2 + \cdots + (a_{mn} - L_m)^2 \\ &< \left(\frac{\varepsilon}{\sqrt{m}}\right)^2 + \left(\frac{\varepsilon}{\sqrt{m}}\right)^2 + \cdots + \left(\frac{\varepsilon}{\sqrt{m}}\right)^2 = \frac{\varepsilon^2}{m} + \frac{\varepsilon^2}{m} + \cdots + \frac{\varepsilon^2}{m} = \varepsilon^2.\end{aligned}$$

Taking square roots of both sides shows that for all $n > N$, $|a_n - L| < \varepsilon$, which shows that $a_n \to L$. □

Example 3.9. Let us apply this theorem to $\mathbb{C}$ (which remember is just $\mathbb{R}^2$ with a special multiplication). Let $c_n = (a_n, b_n) = a_n + ib_n$ be a complex sequence (here we switch notation from a_n to c_n in the theorem and we let $c_{1n} = a_n$ and $c_{2n} = b_n$). Then it follows that $c_n \to c = a + ib$ if and only if $a_n \to a$ and $b_n \to b$. In other words, $c_n \to c$ if and only if the real and imaginary parts of c_n converge to the real and imaginary parts, respectively, of c. For example, from our examples in the previous section, it follows that for any real $a > 0$, we have

$$\frac{1}{n} + ia^{1/n} \to 0 + i \cdot 1 = i.$$

We now prove the fundamental fact that if a sequence converges, then it must be **bounded**. In other words, if $\{a_n\}$ is a convergent sequence in $\mathbb{R}^m$, then there is a constant C such that $|a_n| \leq C$ for all n.

Example 3.10. The sequence $\{n\} = \{1, 2, 3, 4, 5, \ldots\}$ is not bounded by the Archimedean property of $\mathbb{R}$. Also if $a > 1$ is a real number, then the sequence $\{a^n\} = \{a^1, a^2, a^3, \ldots\}$ is not bounded. One way to see this uses Bernoulli's inequality: We can write $a = 1 + r$ where $r > 0$, so by Bernoulli's inequality,

$$a^n = (1 + r)^n \geq 1 + n\,r > n\,r,$$

and $n\,r$ can be made greater than any constant C by the Archimedean property of $\mathbb{R}$. Thus, $\{a^n\}$ cannot be bounded.

Theorem 3.5. *Every convergent sequence is bounded.*

Proof. If $a_n \to L$ in $\mathbb{R}^m$, then with $\varepsilon = 1$ in the definition of convergence, there is an N such that for all $n > N$, we have $|a_n - L| < 1$, which, by the triangle inequality, implies that

$$n > N \quad \Longrightarrow \quad |a_n| = |(a_n - L) + L| \leq |a_n - L| + |L| < 1 + |L|. \tag{3.24}$$

Let k be any natural number greater than N and let

$$C := \max\big\{|a_1|, |a_2|, \ldots, |a_{k-1}|, |a_k|, 1 + |L|\big\}.$$

Then $|a_n| \leq C$ for $n = 1, 2, \ldots, k$, and by (3.24), $|a_n| < C$ for $n > k$. Thus, $|a_n| \leq C$ for all n and hence $\{a_n\}$ is bounded. □

Forming the contrapositive, we know that if a sequence is not bounded, then the sequence cannot converge. Therefore, this theorem can be used to prove that certain sequences *do not* converge.

Example 3.11. Each of following sequences: $\{n\}$, $\{1 + i\,n^2\}$, $\{2^n + i/n\}$ are not bounded, and therefore do not converge.

3.2.2. Real sequences and preservation of inequalities. Real sequences have certain properties that general sequences in $\mathbb{R}^m$ and complex sequences do not have, namely those corresponding to the order properties of $\mathbb{R}$. The order properties of sequences are based on the following lemma.

LEMMA 3.6. *A real sequence $\{a_n\}$ converges to $L \in \mathbb{R}$ if and only if, for every $\varepsilon > 0$ there is an N such that*

$$n > N \quad \Longrightarrow \quad L - \varepsilon < a_n < L + \varepsilon.$$

PROOF. By our interpretation of limits in terms of open balls, we know that $a_n \to L$ just means that given any $\varepsilon > 0$ there is an N such that $n > N \Longrightarrow a_n \in B_\varepsilon(L) = (L - \varepsilon, L + \varepsilon)$. Thus, for $n > N$, we have $L - \varepsilon < a_n < L + \varepsilon$. We can also prove this theorem directly from the original definition of convergence: $a_n \to L$ means that given any $\varepsilon > 0$ there is an N such that

$$n > N \quad \Longrightarrow \quad |a_n - L| < \varepsilon.$$

By our properties of absolute value, $|a_n - L| < \varepsilon$ is equivalent to $-\varepsilon < a_n - L < \varepsilon$, which is equivalent to $L - \varepsilon < a_n < L + \varepsilon$. □

The following theorem is the well-known squeeze theorem. Recall from the beginning part of Section 3.1 that the phrase "for n sufficiently large" means "there is an N such that for $n > N$".

THEOREM 3.7 (**Squeeze theorem**). *Let $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ be sequences in $\mathbb{R}$ with $\{a_n\}$ and $\{c_n\}$ convergent, such that $\lim a_n = \lim c_n$ and for n sufficiently large, $a_n \le b_n \le c_n$. Then the sequence $\{b_n\}$ is also convergent, and*

$$\lim a_n = \lim b_n = \lim c_n.$$

PROOF. Let $L = \lim a_n = \lim c_n$ and let $\varepsilon > 0$. By the tails theorem, we may assume that $a_n \le b_n \le c_n$ for all n. Since $a_n \to L$ there is an N_1 such that for $n > N_1$, $L - \varepsilon < a_n < L + \varepsilon$, and since $c_n \to L$ there is an N_2 such that for $n > N_2$, $L - \varepsilon < c_n < L + \varepsilon$. Let N be the larger of N_1 and N_2. Then for $n > N$,

$$L - \varepsilon < a_n \le b_n \le c_n < L + \varepsilon,$$

which implies that for $n > N$, $L - \varepsilon < b_n < L + \varepsilon$. Thus, $b_n \to L$. □

Example 3.12. Here's a neat sequence involving the squeeze theorem. Consider $\{b_n\}$, where

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

Observe that for $k = 1, 2, \ldots n$, we have

$$0 \le \frac{1}{(n+k)^2} \le \frac{1}{(n+0)^2} = \frac{1}{n^2}.$$

Thus,

$$0 \le \sum_{k=1}^{n} \frac{1}{(n+k)^2} \le \sum_{k=1}^{n} \frac{1}{n^2} = \left(\sum_{k=1}^{n} 1 \right) \cdot \frac{1}{n^2} = n \cdot \frac{1}{n^2}.$$

Therefore,

$$0 \le b_n \le \frac{1}{n}.$$

Since $a_n = 0 \to 0$ and $c_n = 1/n \to 0$, by the squeeze theorem, $b_n \to 0$ as well.

Example 3.13. (**Real numbers as limits of (ir)rational numbers**) We claim that given any $c \in \mathbb{R}$ there are sequences of rational numbers $\{r_n\}$ and irrational numbers $\{q_n\}$, both converging to c. Indeed, for each $n \in \mathbb{N}$ we have $c - \frac{1}{n} < c$, so by Theorem 2.37 there is a rational number r_n and irrational number q_n such that

$$c - \frac{1}{n} < r_n < c \quad \text{and} \quad c - \frac{1}{n} < q_n < c.$$

Since $c - \frac{1}{n} \to c$ and $c \to c$, by the squeeze theorem, we have $r_n \to c$ and $q_n \to c$.

The following theorem states that real sequences preserve inequalities.

Theorem 3.8 (**Preservation of inequalities**). *Let $\{a_n\}$ converge in $\mathbb{R}$.*

(1) If $\{b_n\}$ is convergent and $a_n \le b_n$ for n sufficiently large, then $\lim a_n \le \lim b_n$.
(2) If $c \le a_n \le d$ for n sufficiently large, then $c \le \lim a_n \le d$.

Proof. Since $a_n \to a := \lim a_n$ and $b_n \to b := \lim b_n$ given any $\varepsilon > 0$ there is an N such that for all $n > N$,

$$a - \frac{\varepsilon}{2} < a_n < a + \frac{\varepsilon}{2} \quad \text{and} \quad b - \frac{\varepsilon}{2} < b_n < b + \frac{\varepsilon}{2}$$

By the tails theorem, we may assume that $a_n \le b_n$ for all n. Thus, for $n > N$,

$$a - \frac{\varepsilon}{2} < a_n \le b_n < b + \frac{\varepsilon}{2} \quad \Longrightarrow \quad a - \frac{\varepsilon}{2} < b + \frac{\varepsilon}{2} \quad \Longrightarrow \quad a - b < \varepsilon.$$

By the ε-principle, we have $a - b \le 0$ or $a \le b$, and our first result is proved.

(2) follows from *(1)* applied to the constant sequences $\{c, c, c, \ldots\}$, which converges to c, and $\{d, d, d, \ldots\}$, which converges to d:

$$c = \lim c_n \le \lim a_n \le \lim d_n = d.$$

□

If $c < a_n < d$ for n sufficiently large, must it be true that $c < \lim a_n < d$? The answer is no ... can you give a counterexample?

3.2.3. Subsequences. For the rest of this section we focus on general sequences in $\mathbb{R}^m$ and not just $\mathbb{R}$. A subsequence is just a sequence formed by picking out certain (countably many) terms of a given sequence. More precisely, let $\{a_n\}$ be a sequence in $\mathbb{R}^m$. Let $\nu_1 < \nu_2 < \nu_3 < \cdots$ be a sequence of natural numbers that is increasing. Then the sequence $\{a_{\nu_n}\}$ given by

$$a_{\nu_1},\ a_{\nu_2},\ a_{\nu_3},\ a_{\nu_4},\ \ldots$$

is called a **subsequence** of $\{a_n\}$.

Example 3.14. Consider the sequence

$$\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots, a_n = \frac{1}{n}, \ldots.$$

Choosing $2, 4, 6, \ldots, \nu_n = 2n, \ldots$, we get the subsequence

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \ldots, a_{\nu_n} = \frac{1}{2n}, \ldots.$$

Example 3.15. As another example, choosing $1!, 2!, 3!, 4!, \ldots, \nu_n = n!, \ldots$, we get the subsequence

$$\frac{1}{1!}, \frac{1}{2!}, \frac{1}{3!}, \ldots, a_{\nu_n} = \frac{1}{n!}, \ldots.$$

Notice that both subsequences, $\{1/(2n)\}$ and $\{1/n!\}$ also converge to zero, the same limit as the original sequence $\{1/n\}$. This is a general fact: If a sequence converges, then any subsequence of it must converge to the same limit.

THEOREM 3.9. *Every subsequence of a convergent sequence converges to the same limit as the original sequence.*

PROOF. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$ converging to $L \in \mathbb{R}^m$. Let $\{a_{\nu_n}\}$ be any subsequence and let $\varepsilon > 0$. Since $a_n \to L$ there is an N such that for all $n > N$, $|a_n - L| < \varepsilon$. Since $\nu_1 < \nu_2 < \nu_3 < \ldots$ is an increasing sequence of natural numbers, one can check (for instance, by induction) that $n \leq \nu_n$ for all n. Thus, for $n > N$, we have $\nu_n > N$ and hence for $n > N$, we have $|a_{\nu_n} - L| < \varepsilon$. This proves that $a_{\nu_n} \to L$ and completes the proof. □

This theorem gives perhaps the easiest way to prove that a sequence *does not* converge.

Example 3.16. Consider the sequence

$$i,\ i^2 = -1,\ i^3 = -i,\ i^4 = 1,\ i^5 = i,\ i^6 = -1, \ldots, a_n = i^n, \ldots.$$

Choosing $1, 5, 9, 13, \ldots, \nu_n = 4n - 3, \ldots$, we get the subsequence

$$i,\ i,\ i, i,\ \ldots,$$

which converges to i. On the other hand, choosing $2, 6, 10, 14, \ldots, \nu_n = 4n - 2, \ldots$, we get the subsequence

$$-1,\ -1,\ -1,\ -1, \ldots,$$

which converges to -1. Since these two subsequences do not converge to the same limit, the original sequence $\{i^n\}$ cannot converge. Indeed, if $\{i^n\}$ did converge, then every subsequence of $\{i^n\}$ would have to converge to the same limit as $\{i^n\}$, but we found subsequences that converge to different limits.

3.2.4. Algebra of limits. Let $\{a_n\}$ and $\{b_n\}$ be sequences in $\mathbb{R}^m$. Given any real numbers c, d, we define the **linear combination** of these sequences by c and d as the sequence $\{c\,a_n + d\,b_n\}$. As special case, the sum of these sequences is just the sequence $\{a_n + b_n\}$ and the difference is just the sequence $\{a_n - b_n\}$, and choosing $d = 0$, the multiple of $\{a_n\}$ by c is just the sequence $\{c\,a_n\}$. The **sequence of norms** of the sequence $\{a_n\}$ is the sequence of real numbers $\{|a_n|\}$.

THEOREM 3.10. *Linear combinations and norms of convergent sequences converge to the corresponding linear combinations and norms of the limits.*

PROOF. Consider first the linear combination sequence $\{c\,a_n + d\,b_n\}$. If $a_n \to a$ and $b_n \to b$, we shall prove that $c\,a_n + d\,b_n \to c\,a + d\,b$. Let $\varepsilon > 0$. We need to prove that there is a real number N such that

$$n > N \quad \Longrightarrow \quad |c\,a_n + d\,b_n - (c\,a + d\,b)| < \varepsilon.$$

By the triangle inequality,

$$\begin{aligned} |c\,a_n + d\,b_n - (c\,a + d\,b)| &= |c\,(a_n - a) + d\,(b_n - b)| \\ &\leq |c|\,|a_n - a| + |d|\,|b_n - b|. \end{aligned}$$

Now, since $a_n \to a$, there is an N_1 such that for all $n > N_1$, $|c|\,|a_n - a| < \varepsilon/2$. (If $|c| = 0$, any N_1 will work; if $|c| > 0$, then choose N_1 corresponding to the error $\varepsilon/(2|c|)$ in the definition of convergence for $a_n \to a$.) Similarly, since $b_n \to b$, there is an N_2 such that for all $n > N_2$, $|d|\,|b_n - b| < \varepsilon/2$. Then setting N as the larger of N_1 and N_2, it follows that for $n > N$,

$$|c\,a_n + d\,b_n - (c\,a + d\,b)| \le |c|\,|a_n - a| + |d|\,|b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves that $c\,a_n + d\,b_n \to c\,a + d\,b$.

Assuming that $a_n \to a$, we show that $|a_n| \to |a|$. Let $\varepsilon > 0$. Then there is an N such that $|a_n - a| < \varepsilon$ for all $n > N$. Hence, as a consequence of the triangle inequality (see Property *(iv)* in Theorem 2.41), for $n > N$, we have

$$|\,|a_n| - |a|\,| \le |a_n - a| < \varepsilon,$$

which shows that $|a_n| \to |a|$. □

Let $\{a_n\}$ and $\{b_n\}$ be complex sequences. Given any complex numbers c, d, the same proof detailed above shows that $c\,a_n + d\,b_n \to c\,a + d\,b$. However, being complex sequences we can also multiply these sequences, term by term, defining the **product** sequence as the sequence $\{a_n\,b_n\}$. Also, assuming that $b_n \neq 0$ for each n, we can divide the sequences, term by term, defining the **quotient** sequence as the sequence $\{a_n/b_n\}$.

THEOREM 3.11. *Products of convergent complex sequences converge to the corresponding products of the limits. Quotients of convergent complex sequences, where the denominator sequence is a nonzero sequence converging to a nonzero limit, converge to the corresponding quotient of the limits.*

PROOF. Let $a_n \to a$ and $b_n \to b$. We first prove that $a_n\,b_n \to a\,b$. Let $\varepsilon > 0$. We need to prove that there is a real number N such that for all $n > N$,

$$|a_n\,b_n - a\,b| < \varepsilon.$$

By the triangle inequality,

$$|a_n\,b_n - a\,b| = |a_n(b_n - b) + b(a_n - a)| \le |a_n|\,|b_n - b| + |b|\,|a_n - a|.$$

Since $a_n \to a$, there is an N_1 such that for all $n > N_1$, $|b|\,|a_n - a| < \varepsilon/2$. By Theorem 3.5 there is a constant C such that $|a_n| \le C$ for all n. Since $b_n \to b$, there is an N_2 such that for all $n > N_2$, $C\,|b_n - b| < \varepsilon/2$. Setting N as the larger of N_1 and N_2, it follows that for $n > N$,

$$|a_n\,b_n - a\,b| \le |a_n|\,|b_n - b| + |b|\,|a_n - a| \le C\,|b_n - b| + |b|\,|a_n - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves that $a_n\,b_n \to a\,b$.

We now prove the second statement. If $b_n \neq 0$ for each n and $b \neq 0$, then we shall prove that $a_n/b_n \to a/b$. We can write this limit statement as a product: $a_n \cdot b_n^{-1} \to a \cdot b^{-1}$, so all we have to do is show that $b_n^{-1} \to b^{-1}$. Let $\varepsilon > 0$. We need to prove that there is a real number N such that for all $n > N$,

$$|b_n^{-1} - b^{-1}| = |b_n\,b|^{-1}\,|b_n - b| < \varepsilon.$$

To do so, let N_1 be chosen in accordance with the error $|b|/2$ in the definition of convergence for $b_n \to b$. Then for $n > N_1$,

$$|b| = |b - b_n + b_n| \le |b - b_n| + |b_n| < \frac{|b|}{2} + |b_n|.$$

Bringing $|b|/2$ to the left, for $n > N_1$ we have $|b|/2 < |b_n|$, or

$$|b_n|^{-1} < 2|b|^{-1}, \qquad n > N_1.$$

Now let N_2 be chosen in accordance with the error $|b|^2\varepsilon/2$ in the definition of convergence for $b_n \to b$. Setting N as the larger of N_1 and N_2, it follows that for $n > N$,

$$|b_n^{-1} - b^{-1}| = |b_n\, b|^{-1}\, |b_n - b| < (2|b|^{-1})\, (|b|^{-1})\, |b_n - b| < (2|b|^{-2}) \left(|b|^2 \frac{\varepsilon}{2}\right) = \varepsilon.$$

Thus, $b_n^{-1} \to b^{-1}$ and our proof is complete. □

These two "algebra of limits" theorems can be used to evaluate limits in an easy manner.

Example 3.17. For example, since $\lim \frac{1}{n} = 0$, by our product theorem (Theorem 3.11), we have

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

Example 3.18. In particular, since the constant sequence 1 converges to 1, by our linear combination theorem (Theorem 3.10), for any number a, we have

$$\lim \left(1 + \frac{a}{n^2}\right) = \lim 1 + a \cdot \lim \frac{1}{n^2} = 1 + a \cdot 0 = 1.$$

Example 3.19. Now dividing the top and bottom of $\frac{n^2+3}{n^2+7}$ by $1/n^2$ and using our theorem on quotients and the limit we just found, we obtain

$$\lim \frac{n^2+3}{n^2+7} = \frac{\lim \left(1 + \frac{3}{n^2}\right)}{\lim \left(1 + \frac{7}{n^2}\right)} = \frac{1}{1} = 1.$$

3.2.5. Properly divergent sequences. When dealing with sequences of real numbers, inevitably infinities occur. For instance, we know that the sequence $\{n^2\}$ diverges since it is unbounded. However, in elementary calculus, we would usually write $n^2 \to +\infty$, which suggests that this sequence converges to the number "infinity". We now make this notion precise.

A sequence $\{a_n\}$ of real numbers **diverges to** $+\infty$ if given any real number $M > 0$, there is a real number N such that for all $n > N$, $a_n > M$. The sequence **diverges to** $-\infty$, if for any real number $M < 0$, there is a real number N such that for all $n > N$, $a_n < M$. In the first case we write $\lim a_n = +\infty$ or $a_n \to +\infty$ (sometimes we drop the "+" in front of ∞) and in the second case we write $\lim a_n = -\infty$ or $a_n \to -\infty$. In either case we say that $\{a_n\}$ is **properly divergent**. It is important to understand that the symbols $+\infty$ and $-\infty$ are simply notation and they do not represent real numbers[3]. We now present some examples.

Example 3.20. First we show that for any natural number k, $n^k \to +\infty$. To see this, let $M > 0$. Then we want to prove there is an N such that for all $n > N$, $n^k > M$. To do so, observe that $n^k > M$ if and only if $n > M^{1/k}$. For this reason, we choose $N = M^{1/k}$. With this choice of N, for all $n > N$, we certainly have $n^k > M$ and our proof is complete. Using a very similar argument, one can show that $-n^k \to -\infty$.

[3] It turns out that $\pm\infty$ form part of a number system called the **extended real numbers**, which consists of the real numbers together with the symbols $+\infty = \infty$ and $-\infty$. One can define addition, multiplication, and order in this system, with the exception that subtraction of infinities is not allowed. If you take measure theory, you will study this system.

Example 3.21. In Example 3.10, we showed that given any real number $a > 1$, the sequence $\{a^n\}$ diverges to $+\infty$.

Because $\pm\infty$ are not real numbers, some of the limit theorems we have proved in this section are not valid when $\pm\infty$ are the limits, but many do hold under certain conditions. For example, if $a_n \to +\infty$ and $b_n \to +\infty$, then for any nonnegative real numbers c, d, at least one of which is positive, the reader can check that

$$c\,a_n + d\,b_n \to +\infty.$$

If c, d are nonpositive with at least one of them negative, then $c\,a_n + d\,b_n \to -\infty$. If c and d have opposite signs, then there is no general result. For example, if $a_n = n$, $b_n = n^2$, and $c_n = n + (-1)^n$, then $a_n, b_n, c_n \to +\infty$, but

$$\lim(a_n - b_n) = -\infty, \quad \lim(b_n - c_n) = +\infty, \quad \text{and} \quad \lim(a_n - c_n) \text{ does not exist!}$$

We encourage the reader to think about which limit theorems extend to the case of infinite limits. For example, here is a squeeze law: If $a_n \leq b_n$ for all n sufficiently large and $a_n \to +\infty$, then $b_n \to +\infty$ as well. Some more limit theorems for infinite limits are presented in the exercises (see e.g. Problem 10).

Exercises 3.2.

1. Evaluate the following limits by using limits already proven (in the text or exercises) and invoking the "algebra of limits".
$$(a)\ \lim \frac{(-1)^n\, n}{n^2+5}, \quad (b)\ \lim \frac{(-1)^n}{n+10}, \quad (c)\ \lim \frac{2^n}{3^n+10}, \quad (d)\ \lim \left(7 + \frac{3}{n}\right)^2.$$
2. Why do the following sequences diverge?
$$(a)\ \{(-1)^n\}, \quad (b)\ \left\{a_n = \sum_{k=0}^{n} (-1)^k\right\}, \quad (c)\ \left\{a_n = 2^{n(-1)^n}\right\}, \quad (d)\ \{i^n + 1/n\}.$$
3. Let
$$a_n = \sum_{k=1}^{n} \frac{1}{\sqrt{n^2+k}}\ , \quad b_n = \sum_{k=1}^{n} \frac{1}{\sqrt{n+k}}\ , \quad c_n = \sum_{k=1}^{n} \frac{1}{n^n + n!}.$$
Find $\lim a_n$, $\lim b_n$, and $\lim c_n$.
4. (a) Let $a_1 \in \mathbb{R}$ and for $n \geq 1$, define $a_{n+1} = \frac{\operatorname{sgn}(a_n)+10(-1)^n}{\sqrt{n}}$. Here, $\operatorname{sgn}(x) := 1$ if $x > 0$, $\operatorname{sgn}(x) := 0$ if $x = 0$, and $\operatorname{sgn}(x) := -1$ if $x < 0$. Find $\lim a_n$.
(b) Let $a_1 \in [-1, 1]$ and for $n \geq 1$, define $a_{n+1} = \frac{a_n}{(|a_n|+1)}$. Find $\lim a_n$. Suggestion: Can you prove that $-1/n \leq a_n \leq 1/n$ for all $n \in \mathbb{N}$?
5. If $\{a_n\}$ and $\{b_n\}$ are complex sequences with $\{a_n\}$ bounded and $b_n \to 0$, prove that $a_n\, b_n \to 0$. Why is Theorem 3.11 not applicable in this situation?
6. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$.
(a) Let $b \in \mathbb{R}^m$ and suppose that there is a sequence $\{b_n\}$ in $\mathbb{R}$ with $b_n \to 0$ and $|a_n - b| \leq C|b_n|$ for some $C > 0$ and for all n. Prove that $a_n \to b$.
(b) If $\lim |a_n| = 0$, show that $a_n \to 0$. It is important that zero is the limit in the hypothesis. Indeed, give an example of a sequence in $\mathbb{R}$ for which $\lim |a_n|$ exists and is nonzero, but $\lim a_n$ does not exist.
7. (**The root test for sequences**) Let $\{a_n\}$ be a sequence of positive real numbers with $L := \lim a_n^{1/n} < 1$. (That is, $\{a_n^{1/n}\}$ converges with limit less than 1.)
(i) Show that there is a real number r with $0 < r < 1$ such that $0 < a_n < r^n$ for all n sufficiently large, that is, there is an N such that $0 < a_n < r^n$ for all $n > N$.
(ii) Prove that $\lim a_n = 0$.
(iii) If, however, $L > 1$, prove that a_n is not a bounded sequence, and hence diverges.

(iv) When $L = 1$ the test is inconclusive: Give an example of a convergent sequence and of a divergent sequence, both of which satisfy $L = 1$.

8. (**The ratio test for sequences**) Let $\{a_n\}$ be a sequence of positive real numbers with $L := \lim(a_{n+1}/a_n) < 1$.
   (i) Show that there are real numbers C, r with $C > 0$ and $0 < r < 1$ such that $0 < a_n < C\,r^n$ for all n sufficiently large.
   (ii) Prove that $\lim a_n = 0$.
   (iii) If, however, $L > 1$, prove that a_n is not a bounded sequence, and hence diverges.
   (iv) When $L = 1$ the test is inconclusive: Give an example of a convergent sequence and of a divergent sequence, both of which satisfy $L = 1$.
9. Which of the following sequences are properly divergent? Prove your answers.

$$(a)\ \{\sqrt{n^2+1}\}, \quad (b)\ \{n(-1)^n\}, \quad (c)\ \left\{\frac{3^n-10}{2^n}\right\}, \quad (d)\ \left\{\frac{n}{\sqrt{n+10}}\right\}.$$

10. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers with $\lim a_n = +\infty$ and $b_n \neq 0$ for n large and suppose that for some real number L, we have

$$\lim \frac{a_n}{b_n} = L.$$

   (a) If $L > 0$, prove that $\lim b_n = +\infty$.
   (b) If $L < 0$, prove that $\lim b_n = -\infty$.
   (c) Can you make any conclusions if $L = 0$?

3.3. The monotone criteria, the Bolzano-Weierstrass theorem, and e

In many examples, we proved the convergence of a sequence by first exhibiting an a priori limit of the sequence and then proving that the sequence converged to the exhibited value. For instance, we showed that the sequence $\{1/(n+1)\}$ converges by showing that it converges to 0. Can we still determine if a given sequence converges without producing an a priori limit value? There are two ways to do this, one is called the monotone criterion and the other is the Cauchy criterion. We study the monotone criterion in this section and save Cauchy's criterion for the next. In this section we work strictly with *real numbers*.

3.3.1. Monotone criterion.

A **monotone sequence** $\{a_n\}$ of real numbers is a sequence that is either **nondecreasing**, $a_n \leq a_{n+1}$ for each n:

$$a_1 \leq a_2 \leq a_3 \leq \cdots,$$

or **nonincreasing**, $a_n \geq a_{n+1}$ for each n:

$$a_1 \geq a_2 \geq a_3 \geq \cdots.$$

Example 3.22. Consider the sequence of real numbers $\{a_n\}$ defined inductively as follows:

$$a_1 = 0, \qquad a_{n+1} = \sqrt{1+a_n}, \quad n \in \mathbb{N}. \tag{3.25}$$

Thus,

$$a_1 = 0, \quad a_2 = 1, \quad a_3 = \sqrt{1+\sqrt{1}}, \quad a_4 = \sqrt{1+\sqrt{1+\sqrt{1}}}, \ldots. \tag{3.26}$$

We claim that this sequence is nondecreasing: $0 = a_1 \leq a_2 \leq a_3 \leq \cdots$. To see this, we use induction to prove that $0 \leq a_n \leq a_{n+1}$ for each n. If $n = 1$, then

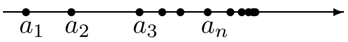


FIGURE 3.2. Bounded monotone (e.g. nondecreasing) sequences must eventually "bunch up" at a point.

$a_1 = 0 \leq 1 = a_2$. Assume that $0 \leq a_n \leq a_{n+1}$. Then $1 + a_n \leq 1 + a_{n+1}$, so using that square roots preserve inequalities (our well-known root rules) we see that

$$a_{n+1} = \sqrt{1 + a_n} \leq \sqrt{1 + a_{n+1}} = a_{n+2}.$$

This establishes the induction step, so we conclude that our sequence $\{a_n\}$ is nondecreasing. We also claim that $\{a_n\}$ is bounded. Indeed, we shall prove that $a_n \leq 3$ for each n. Again we proceed by induction. First, we have $a_1 = 0 \leq 3$. If $a_n \leq 3$, then by definition of a_{n+1}, we have

$$a_{n+1} = \sqrt{1 + a_n} \leq \sqrt{1 + 3} = \sqrt{4} = 2 \leq 3,$$

which proves that $\{a_n\}$ is bounded by 3.

The monotone criterion states that a monotone sequence of real numbers converges to a real number if and only if the sequence is bounded. Intuitively, the terms in a bounded monotone sequence must accumulate at a certain point, see Figure 3.2. In particular, this implies that our recursive sequence (3.25) converges.

THEOREM 3.12 (**Monotone criterion**). *A monotone sequence of real numbers converges if and only if the sequence is bounded.*

PROOF. If a sequence converges, then we know that it must be bounded. So, we need only prove that a bounded monotone sequence converges. So, let $\{a_n\}$ be a bounded monotone sequence. If $\{a_n\}$ is nonincreasing, $a_1 \geq a_2 \geq a_3 \geq \cdots$, then the sequence $\{-a_n\}$ is nondecreasing: $-a_1 \leq -a_2 \leq \cdots$. Thus, if we prove that bounded nondecreasing sequences converge, then $\lim -a_n$ would exist. This would imply that $\lim a_n = -\lim -a_n$ exists too. So it remains to prove our theorem under the assumption that $\{a_n\}$ is nondecreasing: $a_1 \leq a_2 \leq \cdots$. Let $L := \sup\{a_1, a_2, a_3, \ldots\}$, which exists since the sequence is bounded. Let $\varepsilon > 0$. Then $L - \varepsilon$ is smaller than L. Since L is the least upper bound of the set $\{a_1, a_2, a_3, \ldots\}$ and $L - \varepsilon < L$, there must exist an N so that $L - \varepsilon < a_N \leq L$. Since $a_1 \leq a_2 \leq \cdots$, for all $n > N$ we must have $L - \varepsilon < a_n \leq L$ and since $L < L + \varepsilon$, we conclude that

$$n > N \quad \Longrightarrow \quad L - \varepsilon < a_n < L + \varepsilon.$$

Hence, $\lim a_n = L$. □

Example 3.23. The monotone criterion implies that our sequence (3.25) converges, say $a_n \to L \in \mathbb{R}$. Squaring both sides of a_{n+1}, we see that

$$a_{n+1}^2 = 1 + a_n.$$

The subsequence $\{a_{n+1}\}$ also converges to L by Theorem 3.9, therefore by the algebra of limits,

$$L^2 = \lim a_{n+1}^2 = \lim(1 + a_n) = 1 + L,$$

or solving for L, we get (using the quadratic formula),

$$L = \frac{1 \pm \sqrt{5}}{2}.$$

Since $0 = a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n \to L$, and limits preserve inequalities (Theorem 3.8), the limit L cannot be negative, so we conclude that

$$L = \frac{1+\sqrt{5}}{2},$$

the number called the **golden ratio** and is denoted by Φ. In view of the expressions found in (3.26), we can interpret Φ as the infinite "continued square root":

$$\boxed{\Phi = \frac{1+\sqrt{5}}{2} = \sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}}}}. } \tag{3.27}$$

There are many stories about Φ; unfortunately, many of them are false, see [**146**].

Our next important theorem is the monotone subsequence theorem. There are many nice proofs of this theorem, cf. the articles [**158**] [**20**], [**223**]. Given any set A of real numbers, a number a is said to be the **maximum** of A if $a \in A$ and $a = \sup A$, in which case we write $a = \max A$.

THEOREM 3.13 (**Monotone subsequence theorem**). *Any sequence of real numbers has a monotone subsequence.*

PROOF. Let $\{a_n\}$ be a sequence of real numbers. Then the statement "for every $n \in \mathbb{N}$, the maximum of the set $\{a_n, a_{n+1}, a_{n+2}, \ldots\}$ exists" is either a true statement, or it's false, which means "there is an $m \in \mathbb{N}$ such that the maximum of the set $\{a_m, a_{m+1}, a_{m+2}, \ldots\}$ does not exist."

Case 1: Suppose that we are in the first case: for each n, $\{a_n, a_{n+1}, a_{n+2}, \ldots\}$ has a greatest member. In particular, we can choose a_{ν_1} such that

$$a_{\nu_1} = \max\{a_1, a_2, \ldots\}.$$

Now $\{a_{\nu_1+1}, a_{\nu_1+2}, \ldots\}$ has a greatest member, so we can choose a_{ν_2} such that

$$a_{\nu_2} = \max\{a_{\nu_1+1}, a_{\nu_1+2}, \ldots\}.$$

Since a_{ν_2} is obtained by taking the maximum of a smaller set of elements, we have $a_{\nu_2} \leq a_{\nu_1}$. Let

$$a_{\nu_3} = \max\{a_{\nu_2+1}, a_{\nu_2+2}, \ldots\}.$$

Since a_{ν_3} is obtained by taking the maximum of a smaller set of elements than the set defining a_{ν_2}, we have $a_{\nu_3} \leq a_{\nu_2}$. Continuing by induction we construct a monotone (nonincreasing) subsequence.

Case 2: Suppose that the maximum of the set $A = \{a_m, a_{m+1}, a_{m+2}, \ldots\}$ does not exist, where $m \geq 1$. Let $a_{\nu_1} = a_m$. Since A has no maximum, there is a $\nu_2 > m$ such that

$$a_m < a_{\nu_2},$$

for, if there were no such a_{ν_2}, then a_m would be a maximum element of A, which we know is not possible. Since none of the elements $a_m, a_{m+1}, \ldots, a_{\nu_2}$ is a maximum element of A, there must exist an $\nu_3 > \nu_2$ such that

$$a_{\nu_2} < a_{\nu_3},$$

for, otherwise one of $a_m, \dots, a_{n_2}$ would be a maximum element of A. Similarly, since none of $a_m, \dots, a_{\nu_2}, \dots, a_{\nu_3}$ is a maximum element of A, there must exist an $\nu_4 > \nu_3$ such that

$$a_{\nu_3} < a_{\nu_4}.$$

Continuing by induction we construct a monotone (nondecreasing) sequence $\{a_{\nu_k}\}$. □

3.3.2. The Bolzano-Weierstrass theorem. The following theorem named after Bernard Bolzano (1781–1848) and Karl Weierstrass (1815–1897) is one of the most important results in analysis and will be frequently employed in the sequel.

Theorem 3.14 (**Bolzano-Weierstrass theorem for $\mathbb{R}$**). *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence. In fact, if the sequence is contained in a closed interval I, then the limit of the convergent subsequence is also in I.*

Proof. Let $\{a_n\}$ be a bounded sequence in $\mathbb{R}$. By the monotone subsequence theorem, this sequence has a monotone subsequence $\{a_{\nu_n}\}$, which of course is also bounded. By the monotone criterion (Theorem 3.12), this subsequence converges to some limit value L. Suppose that $\{a_n\}$ is contained in a closed interval $I = [a, b]$. Then $a \le a_{\nu_n} \le b$ for each n. Since limits preserve inequalities, the limit L of the subsequence $\{a_{\nu_n}\}$ also lies in $[a, b]$. □

Using induction on m (we already did the $m = 1$ case), we leave the proof of the following generalization to you, if you're interested.

Theorem 3.15 (**Bolzano-Weierstrass theorem for $\mathbb{R}^m$**). *Every bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

3.3.3. The number e. We now define Euler's constant e by a method that has been around for ages, cf. [**120**, p. 82], [**243**]. Consider the two sequences whose terms are given by

$$a_n = \left(1 + \frac{1}{n}\right)^n = \left(\frac{n+1}{n}\right)^n \quad \text{and} \quad b_n = \left(1 + \frac{1}{n}\right)^{n+1} = \left(\frac{n+1}{n}\right)^{n+1},$$

where $n = 1, 2, \dots$. We shall prove that the sequence $\{a_n\}$ is bounded above and is **strictly increasing**, which means that $a_n < a_{n+1}$ for all n. We'll also prove that $\{b_n\}$ is bounded below and **strictly decreasing**, which means that $b_n > b_{n+1}$ for all n. In particular, the limits $\lim a_n$ and $\lim b_n$ exist by the monotone criterion. Notice that

$$b_n = a_n \left(1 + \frac{1}{n}\right),$$

and $1 + 1/n \to 1$, so if sequences $\{a_n\}$ and $\{b_n\}$ converge, they must converge to the same limit. This limit is denoted by the letter e, introduced in 1727 by Euler perhaps because "e" is the first letter in "exponential" [**36**, p. 442], *not* because "e" is the first letter of his last name!

The proof that the sequences above are monotone follow from Bernoulli's inequality. First, to see that $b_{n-1} > b_n$ for $n \ge 2$, observe that

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

According to Bernoulli's inequality, we have

$$\left(1+\frac{1}{n^2-1}\right)^n > 1+\frac{n}{n^2-1} > 1+\frac{n}{n^2} = \frac{n+1}{n},$$

which implies that

$$\frac{b_{n-1}}{b_n} > \frac{n+1}{n}\cdot\frac{n}{n+1} = 1 \quad\Longrightarrow\quad b_{n-1} > b_n$$

and proves that $\{b_n\}$ is strictly decreasing. Certainly $b_n > 0$ for each n, so the sequence b_n is bounded below and hence converges.

To see that $a_{n-1} < a_n$ for $n \geq 2$, we proceed in a similar manner:

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

Bernoulli's inequality for $n \geq 2$ implies that

$$\left(1-\frac{1}{n^2}\right)^n > 1-\frac{n}{n^2} = 1-\frac{1}{n} = \frac{n-1}{n},$$

so

$$\frac{a_n}{a_{n-1}} > \frac{n-1}{n}\cdot\frac{n}{n-1} = 1 \quad\Longrightarrow\quad a_{n-1} < a_n.$$

This shows that $\{a_n\}$ is strictly increasing. Finally, since $a_n < b_n < b_1 = 4$, the sequence $\{a_n\}$ is bounded above.

In conclusion, we have proved that the limit

$$\boxed{e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n}$$

exists, which equals by definition the number denoted by e. Moreover, we have also derived the inequality

$$\boxed{\left(1+\frac{1}{n}\right)^n < e < \left(1+\frac{1}{n}\right)^{n+1}, \qquad \text{for all } n.} \tag{3.28}$$

We shall need this inequality later when we discuss Euler-Mascheroni's constant. This inequality is also useful in studying Stirling's formula, which we now describe. Recall that $0! := 1$ and given a positive integer n, we define $n!$ (which we read "n factorial") as $n! := 1\cdot 2\cdot 3\cdots n$. For n positive, observe that $n!$ is less than n^n, or equivalently, $\sqrt[n]{n!}/n < 1$. A natural question to ask is: How much less than one is the ratio $\sqrt[n]{n!}/n$? Using (3.28), in Problem 5 you will prove that

$$\lim \frac{\sqrt[n]{n!}}{n} = \frac{1}{e}, \qquad (\textbf{Stirling's formula}). \tag{3.29}$$

Exercises 3.3.

1. (a) Show that the sequence defined inductively by $a_{n+1} = \frac{1}{3}(2a_n + 4)$ with $a_1 = 0$ is nondecreasing and bounded above by 6. Suggestion: Use induction, e.g. prove that $a_n \leq a_{n+1}$ and $a_n \leq 6$ by induction on n. Determine the limit.

   (b) Let $a_1 = 1$. Show that the sequence defined inductively by $a_{n+1} = \sqrt{2\,a_n}$ is nondecreasing and bounded above by 2. Determine the limit.

(c) Let $a \geq 2$. Show that the sequence defined inductively by $a_{n+1} = 2 - a_n^{-1}$ with $a_1 = a$ is a bounded monotone sequence. Determine the limit. Suggestion: Pick e.g. $a = 2$ or $a = 10$ and calculate a few values of a_n to conjecture if $\{a_n\}$ is, for general $a > 1$, nondecreasing or nonincreasing. Also conjecture what a bound may be from these examples. Now prove your conjecture using induction.

(d) Let $a > 0$. Show that the sequence defined inductively by $a_{n+1} = a_n/(1 + 2a_n)$ with $a_1 = a$ is a bounded monotone sequence. Determine the limit.

(e) Show that the sequence defined inductively by $a_{n+1} = 5 + \sqrt{a_n - 5}$ with $a_1 = 10$ is a bounded monotone sequence. Determine the limit.

(f) Show that the sequence with $a_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$ is a bounded monotone sequence. The limit of this sequence is not at all obvious (it turns out to be $\log 2$, which we'll study in Section 4.6).

2. In this problem we give two different ways to determine square roots using sequences.

(1) Let $a > 0$. Let a_1 be any positive number and define

$$a_{n+1} = \frac{1}{2}\left(a_n + \frac{a}{a_n}\right), \qquad n \geq 1.$$

(a) Show that $a_n > 0$ for all n and $a_{n+1}^2 - a \geq 0$ for all n.

(b) Show that $\{a_n\}$ is nonincreasing for $n \geq 2$.

(c) Conclude that $\{a_n\}$ converges and determine its limit.

(2) Let $0 \leq a \leq 1$. Show that the sequence defined inductively by $a_{n+1} = a_n + \frac{1}{2}(a - a_n^2)$ with $a_1 = 0$ is nondecreasing and bounded above by $\sqrt{a}$. Determine the limit. Suggestion: To prove that a_{n+1} is bounded above by $\sqrt{a}$, assume a_n is and write $a_n = \sqrt{a} - \varepsilon$ where $\varepsilon \geq 0$.

3. (Cf. [**250**]) In this problem we analyze the constant e based on arithmetic-geometric mean inequality (AGMI); see Problem 7 of Exercises 2.2. Recall that the arithmetic-geometric mean states that given any $n+1$ nonnegative real numbers $x_1, \ldots, x_{n+1}$,

$$x_1 \cdot x_2 \cdots x_{n+1} \leq \left(\frac{x_1 + x_2 + \cdots + x_{n+1}}{n+1}\right)^{n+1}.$$

(i) Put $x_k = (1 + 1/n)$ for $k = 1, \ldots, n$ and $x_{n+1} = 1$, in the AGMI to prove that the sequence $a_n = (1 + \frac{1}{n})^n$ is nondecreasing.

(ii) If $b_n = (1 + 1/n)^{n+1}$, then show that for $n \geq 2$,

$$\frac{b_n}{b_{n-1}} = \left(1 - \frac{1}{n^2}\right)^n \left(1 + \frac{1}{n}\right) = \underbrace{\left(1 - \frac{1}{n^2}\right) \cdots \left(1 - \frac{1}{n^2}\right)}_{n \text{ times}} \left(1 + \frac{1}{n}\right).$$

Applying the AGMI to the right hand side, show that $b_n/b_{n-1} \leq 1$, which shows that the sequence $\{b_n\}$ is nonincreasing.

(iii) Conclude that both sequences $\{a_n\}$ and $\{b_n\}$ converge. Of course, just as in the text we denote their common limit by e.

4. (**Continued roots**) For more on this subject, see [**4**], [**101**], [**149**, p. 775], [**215**], [**106**].

(1) Fix $k \in \mathbb{N}$ with $k \geq 2$ and fix $a > 0$. Show that the sequence defined inductively by $a_{n+1} = \sqrt[k]{a + a_n}$ with $a_1 = \sqrt[k]{a}$ is a bounded monotone sequence. Prove that the limit L is a root of the equation $x^k - x - a = 0$. Can you see why L can be thought of as

$$L = \sqrt[k]{a + \sqrt[k]{a + \sqrt[k]{a + \sqrt[k]{a + \cdots}}}}.$$

(2) Let $\{a_n\}$ be a sequence of nonnegative real numbers and for each n, define

$$\alpha_n := \sqrt{a_1 + \sqrt{a_2 + \sqrt{a_3 + \sqrt{\cdots + \sqrt{a_n}}}}}.$$

Prove that $\{\alpha_n\}$ converges if and only if there is a constant $M \geq 0$ such that $\sqrt[2^n]{a_n} \leq M$ for all n. Suggestion: To prove the "only if" portion, prove that $\sqrt[2^n]{a_n} \leq \alpha_n$. To prove the "if", prove that $\alpha_n \leq \sqrt{M + \sqrt{M^{2^2} + \sqrt{\cdots + \sqrt{M^{2^n}}}}} = Mb_n$ where $b_n = \sqrt{1 + \sqrt{1 + \sqrt{\cdots + \sqrt{1}}}}$ is defined in (3.25); in particular, we showed that $b_n \leq 3$. Now setting $a_n = n$ for all n, show that

$$\sqrt{1 + \sqrt{2 + \sqrt{3 + \sqrt{4 + \sqrt{5 + \cdots}}}}}$$

exists. This number is called **Kasner's number** named after Edward Kasner (1878–1955) and is approximately 1.75793....

5. In this problem we prove (3.29).
   (a) Prove that for each natural number n, $(n-1)! \leq n^n e^{-n} e \leq n!$. Suggestion: Can you use induction and (3.28)? (You can also prove these inequalities using integrals as in [**128**, p. 219], but using (3.28) gives an "elementary" proof that is free of integration theory.)
   (b) Using (a), prove that for every natural number n,
   $$\frac{e^{1/n}}{e} \leq \frac{\sqrt[n]{n!}}{n} \leq \frac{e^{1/n}\, n^{1/n}}{e}.$$
   (c) Now prove (3.29). Using (3.29), prove that
   $$\lim \left(\frac{(3n)!}{n^{3n}} \right)^{1/n} = \frac{27}{e^3} \qquad \text{and} \qquad \lim \left(\frac{(3n)!}{n!\, n^{2n}} \right)^{1/n} = \frac{27}{e^2}.$$

3.4. Completeness and the Cauchy criteria for convergence

The monotone criterion gives a criterion for convergence (in $\mathbb{R}$) of a monotone sequence of real numbers. Now what if the sequence is not monotone? The Cauchy criterion, originating with Bolzano, but then made into a formulated "criterion" by Cauchy [**120**, p. 87], gives a convergence criterion for general sequences of real numbers and more generally, sequences of complex numbers and vectors.

3.4.1. Cauchy sequences. A sequence $\{a_n\}$ in $\mathbb{R}^m$ is said to be **Cauchy** if,

$$\boxed{\text{for every } \varepsilon > 0, \text{ there is an } N \in \mathbb{R} \text{ such that } \quad k, n > N \implies |a_k - a_n| < \varepsilon.}$$

Intuitively, all the a_n's get closer together as the indices n gets larger and larger.

Example 3.24. The sequence of real numbers with $a_n = \frac{2n-1}{n-3}$ and $n \geq 4$ is Cauchy. To see this, let $\varepsilon > 0$. We need to prove that there is a real number N such that

$$k, n > N \quad \implies \quad \left| \frac{2k-1}{k-3} - \frac{2n-1}{n-3} \right| < \varepsilon.$$

To see this, we "massage" the right-hand expression:

$$\begin{aligned} \left| \frac{2k-1}{k-3} - \frac{2n-1}{n-3} \right| &= \left| \frac{(2k-1)(n-3) - (2n-1)(k-3)}{(k-3)(n-3)} \right| \\ &= \left| \frac{5(n-k)}{(k-3)(n-3)} \right| \leq \left| \frac{5n}{(k-3)(n-3)} \right| + \left| \frac{5k}{(k-3)(n-3)} \right|. \end{aligned}$$

Now observe that for $n \geq 4$, we have $\frac{n}{4} \geq 1$, so

$$n - 3 \geq n - \left(3 \cdot \frac{n}{4}\right) = n - \frac{3n}{4} = \frac{n}{4} \quad \implies \quad \frac{1}{n-3} \leq \frac{4}{n}.$$

Thus, for $n, k \geq 4$, we have

$$\left|\frac{5n}{(k-3)(n-3)}\right| + \left|\frac{5k}{(k-3)(n-3)}\right| < 5n \cdot \frac{4}{k} \cdot \frac{4}{n} + 5k \cdot \frac{4}{k} \cdot \frac{4}{n} = \frac{80}{k} + \frac{80}{n}.$$

Thus, our "massaged" expression has been fully relaxed as an inequality

$$\text{for } k, n \geq 4, \qquad \left|\frac{2k-1}{k-3} - \frac{2n-1}{n-3}\right| < \frac{80}{k} + \frac{80}{n}.$$

Thus, we can make the left-hand side less than ε by making the right-hand side less than ε, and we can do this by noticing that we can make

$$\frac{80}{k} + \frac{80}{n} < \varepsilon \quad \text{by making} \quad \frac{80}{k}, \frac{80}{n} < \frac{\varepsilon}{2},$$

or after solving for k and n, we must have $k, n > 160/\varepsilon$. For this reason, let us pick N to be the larger of 3 and $160/\varepsilon$. Let $k, n > N$ (that is, $k, n \geq 4$ and $k, n > 160/\varepsilon$). Then,

$$\left|\frac{2k-1}{k-3} - \frac{2n-1}{n-3}\right| < \frac{80}{k} + \frac{80}{n} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This shows that the sequence $\{a_n\}$ is Cauchy. Notice that this sequence, $\{\frac{2n-1}{n-3}\}$, converges (to the number 2).

Example 3.25. Here's a more sophisticated example of a Cauchy sequence. Let $a_1 = 1$, $a_2 = 1/2$, and for $n \geq 2$, we let a_n be the arithmetic mean between the previous two terms:

$$a_n = \frac{a_{n-2} + a_{n-1}}{2}, \qquad n > 2.$$

Thus, $a_1 = 1$, $a_2 = 1/2$, $a_3 = 3/4$, $a_4 = 5/8, \ldots$ so this sequence is certainly not monotone. However, we shall prove that $\{a_n\}$ is Cauchy. To do so, we first prove by induction that

$$a_n - a_{n+1} = \frac{(-1)^{n+1}}{2^n}. \tag{3.30}$$

Since $a_1 = 1$ and $a_2 = 1/2$, this equation holds for $n = 1$. Assume that the equation holds for n. Then

$$a_{n+1} - a_{n+2} = a_{n+1} - \frac{a_{n+1} + a_n}{2} = -\frac{1}{2}\big(a_n - a_{n+1}\big) = -\frac{1}{2}\frac{(-1)^{n+1}}{2^n} = \frac{(-1)^{n+2}}{2^{n+1}},$$

which proves the induction step. With (3.30) at hand, we can now show that the sequence $\{a_n\}$ is Cauchy. Let k, n be any natural numbers. By symmetry we may assume that $k \leq n$ (otherwise just switch k and n in what follows), let us say $n = k + j$ where $j \geq 0$. Then according to (3.30) and the sum of a geometric progression (2.3), we can write

$$\begin{aligned} a_k - a_n &= a_k - a_{k+j} \\ &= \big(a_k - a_{k+1}\big) + \big(a_{k+1} - a_{k+2}\big) + \big(a_{k+2} - a_{k+3}\big) + \cdots + \big(a_{k+j-1} - a_{k+j}\big) \\ &= \frac{(-1)^{k+1}}{2^k} + \frac{(-1)^{k+2}}{2^{k+1}} + \frac{(-1)^{k+3}}{2^{k+3}} + \cdots + \frac{(-1)^{k+j}}{2^{k+j-1}} \\ &= \frac{(-1)^{k+1}}{2^k}\left[1 + \frac{-1}{2} + \Big(\frac{-1}{2}\Big)^2 + \cdots + \Big(\frac{-1}{2}\Big)^{j-1}\right] \\ &= \frac{(-1)^{k+1}}{2^k} \cdot \frac{1 - (-1/2)^j}{1 - (-1/2)} = \frac{(-1)^{k+1}}{2^k} \cdot \frac{2}{3} \cdot \big(1 - (-1/2)^j\big). \qquad (3.31) \end{aligned}$$

Since $\frac{2}{3} \cdot \left|1 - (-1/2)^j\right| \leq \frac{2}{3} \cdot \left(1 + 1\right) \leq 2$, we conclude that

$$|a_k - a_n| \leq \frac{1}{2^{k-1}}, \qquad \text{for all } k, n \text{ with } k \leq n.$$

Now let $\varepsilon > 0$. Since $1/2 < 1$, we know that $1/2^{k-1} = 2 \cdot (1/2)^k \to 0$ as $k \to \infty$ (see Example 3.5 in Subsection 3.1.3). Therefore there is an N such that for all $k > N$, $1/2^{k-1} < \varepsilon$. Let $k, n > N$ and again by symmetry, we may assume that $k \leq n$. In this case, we have

$$|a_k - a_n| \leq \frac{1}{2^{k-1}} < \varepsilon.$$

This proves that the sequence $\{a_n\}$ is Cauchy. Moreover, we claim that this sequence also converges. Indeed, by (3.31) with $k = 1$, so that $n = 1 + j$ or $j = n - 1$, we see that

$$1 - a_n = \frac{1}{2} \cdot \frac{2}{3} \cdot \left(1 - (-1/2)^{n-1}\right) \quad \Longrightarrow \quad a_n = 1 - \frac{1}{3} \cdot \left(1 - (-1/2)^{n-1}\right).$$

Since $|(-1/2)| = 1/2 < 1$, we know that $(-1/2)^{n-1} \to 0$. Taking $n \to \infty$, we conclude that

$$\lim a_n = 1 - \frac{1}{3}\,(1 - 0) = \frac{2}{3}.$$

We have thus far gave examples of two Cauchy sequences and we have observed that both sequences converge. In Theorem 3.17 we shall prove that any Cauchy sequence converges, and conversely, every convergent sequence is also Cauchy.

3.4.2. Cauchy criterion.

The following two proofs use the "$\varepsilon/2$-trick."

LEMMA 3.16. *If a subsequence of a Cauchy sequence in $\mathbb{R}^m$ converges, then the whole sequence converges too, and with the same limit as the subsequence.*

PROOF. Let $\{a_n\}$ be a Cauchy sequence and assume that $a_{\nu_n} \to L$ for some subsequence of $\{a_n\}$. We shall prove that $a_n \to L$. Let $\varepsilon > 0$. Since $\{a_n\}$ is Cauchy, there is an N such that

$$k, n > N \quad \Longrightarrow \quad |a_k - a_n| < \frac{\varepsilon}{2}.$$

Since $a_{\nu_n} \to L$ there is a natural number $k \in \{\nu_1, \nu_2, \nu_3, \nu_4, \ldots\}$ with $k > N$ such that

$$|a_k - L| < \frac{\varepsilon}{2}.$$

Now let $n > N$ be arbitrary. Then using the triangle inequality and the two inequalities we just wrote down, we see that

$$|a_n - L| = |a_n - a_k + a_k - L| \leq |a_n - a_k| + |a_k - L| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves that $a_n \to L$ and our proof is complete. □

THEOREM 3.17 (**Cauchy criterion**). *A sequence in $\mathbb{R}^m$ converges if and only if it is Cauchy.*

PROOF. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$ converging to $L \in \mathbb{R}^m$. We shall prove that the sequence is Cauchy. Let $\varepsilon > 0$. Since $a_n \to L$ there is an N such that for all $n > N$, we have $|a_n - L| < \frac{\varepsilon}{2}$. Hence, by the triangle inequality

$$k, n > N \quad \Longrightarrow \quad |a_k - a_n| \leq |a_k - L| + |L - a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves that a convergent sequence is also Cauchy.

Now let $\{a_n\}$ be Cauchy. We shall prove that this sequence also converges. First we prove that the sequence is bounded. To see this, let us put $\varepsilon = 1$ in the definition of being a Cauchy sequence; then there is an N such that for all $k, n \geq N$, we have $|a_k - a_n| < 1$. Fix any $k > N$. Then by the triangle inequality, for any $n > k$, we have

$$|a_n| = |a_n - a_k + a_k| \leq |a_n - a_k| + |a_k| \leq 1 + |a_k|.$$

Hence, for any natural number n, this inequality implies that

$$|a_n| \leq \max\{|a_1|,\ |a_2|,\ |a_3|, \ldots, |a_{k-1}|,\ 1 + |a_k|\}.$$

This shows that the sequence $\{a_n\}$ is bounded. Therefore, the Bolzano-Weierstrass theorem implies that this sequence has a convergent subsequence. Our lemma now guarantees that the whole sequence $\{a_n\}$ converges. □

Because every Cauchy sequence in $\mathbb{R}^m$ converges in $\mathbb{R}^m$, we say that $\mathbb{R}^m$ is **complete**. This property of $\mathbb{R}^m$ is essential to many objects in analysis, e.g. series, differentiation, integration, ..., all of which use limit processes. $\mathbb{Q}$ is an example of something that is not complete.

Example 3.26. The sequence

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ 1.41421, \ldots$$

is a Cauchy sequence of rational numbers, but its limit (which is supposed to be $\sqrt{2}$) does not exist as a rational number! (By the way, we'll study decimal expansions of real numbers in Section 3.8.)

From this example you can imagine the difficulties the noncompleteness of $\mathbb{Q}$ can cause when trying to do analysis with strictly rational numbers.

3.4.3. Contractive sequences.

Cauchy's criterion is important because it allows us to determine whether a sequence converges or diverges by instead proving that the sequence is or is not Cauchy. Thus, we now focus on how to determine whether or not a sequence is Cauchy. Of course, we could appeal directly to the definition of a Cauchy sequence, but unfortunately it is sometimes difficult to show that a given sequence is Cauchy directly from the definition. However, for a wide variety of applications it is often easier to show that a sequence is "contractive," which automatically implies that it is Cauchy (see the contractive sequence theorem below), and hence by the Cauchy criterion, must converge.

A sequence $\{a_n\}$ in $\mathbb{R}^m$ is a said to be **contractive** if there is a $0 < r < 1$ such that for all n,

$$|a_n - a_{n+1}| \leq r\,|a_{n-1} - a_n|. \tag{3.32}$$

Theorem 3.18 (**Contractive sequence theorem**). *Every contractive sequence converges.*

Proof. Let $\{a_n\}$ be a contractive sequence. Then with $n = 2$ in (3.32), we see that

$$|a_2 - a_3| \leq r\,|a_1 - a_2|,$$

and with $n = 3$,

$$|a_3 - a_4| \leq r\,|a_2 - a_3| \leq r \cdot r\,|a_1 - a_2| = r^2\,|a_1 - a_2|.$$

By induction, for $n \geq 2$ we get

$$|a_n - a_{n+1}| \leq C\, r^{n-1}, \quad \text{where } \; C = |a_1 - a_2|. \tag{3.33}$$

To prove that $\{a_n\}$ converges, all we have to do is prove the sequence is Cauchy. Let $k, n \geq 2$. By symmetry we may assume that $n \leq k$ (otherwise just switch k and n in what follows), say $k = n + j$ where $j \geq 0$. Then according to (3.33), the triangle inequality, and the geometric progression (2.3), we can write

$$\begin{aligned} |a_n - a_k| &= \big|\big(a_n - a_{n+1}\big) + \big(a_{n+1} - a_{n+2}\big) + \cdots + \big(a_{n+j-1} - a_{n+j}\big)\big| \\ &\leq \big|a_n - a_{n+1}\big| + \big|a_{n+1} - a_{n+2}\big| + \big|a_{n+2} - a_{n+3}\big| + \cdots + \big|a_{n+j-1} - a_{n+j}\big| \\ &= C\, r^{n-1} + C\, r^{n} + C\, r^{n+1} + \cdots + C\, r^{n+j-2} \\ &= C\, r^{n-1}\Big[1 + r + r^2 + \cdots + r^{j-1}\Big] = C\, r^{n-1}\frac{1 - r^j}{1 - r} \leq \frac{C}{1 - r}\, r^{n-1}. \end{aligned}$$

We are now ready to prove that the sequence $\{a_n\}$ is Cauchy. Let $\varepsilon > 0$. Since $r < 1$, we know that $\frac{C}{1-r}\, r^{n-1} \to 0$ as $n \to \infty$. Therefore there is an N, which we may assume is greater than 1, such that for all $n > N$, $\frac{C}{1-r}\, r^{n-1} < \varepsilon$. Let $k, n > N$. Then $k, n \geq 2$ and by symmetry, we may assume that $n \leq k$, in which case by the above calculation we find that

$$|a_n - a_k| \leq \frac{C}{1 - r}\, r^{n-1} < \varepsilon.$$

This proves that the sequence $\{a_n\}$ is Cauchy. □

By the tails theorem (Theorem 3.3), a sequence $\{a_n\}$ will converge as long as (3.32) holds for sufficiently n large. We now consider an example.

Example 3.27. Define

$$a_1 = 1 \quad \text{and} \quad a_{n+1} = \sqrt{9 - 2a_n}, \quad n \geq 1. \tag{3.34}$$

It is not a priori obvious that this sequence is well-defined; how do we know that $9 - 2a_n \geq 0$ for all n so that we can take the square root $\sqrt{9 - 2a_n}$ to define a_{n+1}? Thus, we need to show that a_n cannot get "too big" so that $9 - 2a_n$ becomes negative. This is accomplished through the following estimate:

(3.35) For all $n \in \mathbb{N}$, a_n is defined and $0 \leq a_n \leq 3$.

This certainly holds for $n = 1$. If (3.35) holds for a_n, then after manipulating the inequality we see that $3 \leq 9 - 2a_n \leq 9$. In particular, the square root $a_{n+1} = \sqrt{9 - 2a_n}$ is well-defined, and

$$0 \leq \sqrt{3} \leq a_{n+1} = \sqrt{9 - 2a_n} \leq \sqrt{9} = 3,$$

so (3.35) holds for a_{n+1}. Therefore, (3.35) holds for every n. Now multiplying by conjugates, for $n \geq 2$, we obtain

$$\begin{aligned} a_n - a_{n+1} &= \left(\sqrt{9 - 2a_{n-1}} - \sqrt{9 - 2a_n}\right) \frac{\sqrt{9 - 2a_{n-1}} + \sqrt{9 - 2a_n}}{\sqrt{9 - 2a_{n-1}} + \sqrt{9 - 2a_n}} \\ &= \frac{-2a_{n-1} + 2a_n}{\sqrt{9 - 2a_{n-1}} + \sqrt{9 - 2a_n}} \\ &= \frac{2}{\sqrt{9 - 2a_{n-1}} + \sqrt{9 - 2a_n}} \cdot (-a_{n-1} + a_n) \end{aligned}$$

The smallest the denominator can possibly be is when a_{n-1} and a_n are the largest they can be, which according to (3.35), is at most 3. It follows that for any $n \geq 2$,

$$\frac{2}{\sqrt{9-2a_{n-1}}+\sqrt{9-2a_n}} \leq \frac{2}{\sqrt{9-2\cdot 3}+\sqrt{9-2\cdot 3}} = \frac{2}{2\sqrt{3}} = r,$$

where $r := \frac{1}{\sqrt{3}} < 1$. Thus, for any $n \geq 2$,

$$|a_n - a_{n+1}| = \frac{2}{\sqrt{9-2a_{n-1}}+\sqrt{9-2a_n}}\,|a_{n-1}-a_n| \leq r\,|a_{n-1}-a_n|.$$

This proves that the sequence $\{a_n\}$ is contractive and therefore $a_n \to L$ for some real number L. Because $a_n \geq 0$ for all n and limits preserve inequalities we must have $L \geq 0$ too. Moreover, by (3.34), we have

$$L^2 = \lim a_{n+1}^2 = \lim(9-2a_n) = 9-2L,$$

which implies that $L^2+2L-9=0$. Solving this quadratic equation and taking the positive root we conclude that $L=\sqrt{10}-1$.

Exercises 3.4.

1. Prove directly, via the definition, that the following sequences are Cauchy.

$$(a)\ \left\{10+\frac{(-1)^n}{\sqrt{n}}\right\},\quad (b)\ \left\{\left(7+\frac{3}{n}\right)^2\right\},\quad (c)\ \left\{\frac{n^2}{n^2-5}\right\}.$$

2. Negate the statement that a sequence $\{a_n\}$ is Cauchy. With your negation, prove that the following sequences are not Cauchy (and hence cannot converge).

$$(a)\ \{(-1)^n\},\qquad (b)\ \left\{a_n=\sum_{k=0}^{n}(-1)^n\right\},\qquad (c)\ \{i^n+1/n\}.$$

3. Prove that the following sequences are contractive, then determine their limits.
   (a) Let $a_1=0$ and $a_{n+1}=(2a_n-3)/4$.
   (b) Let $a_1=1$ and $a_{n+1}=\frac{1}{5}a_n^2-1$.
   (c) Let $a_1=0$ and $a_{n+1}=\frac{1}{8}a_n^3+\frac{1}{4}a_n+\frac{1}{2}$.
   (d) Let $a_1=1$ and $a_{n+1}=\frac{1}{1+3a_n}$. Suggestion: Prove that $\frac{1}{4}\leq a_n\leq 1$ for all n.
   (e) (Cf. Example 3.27.) Let $a_1=1$ and $a_{n+1}=\sqrt{5-2a_n}$.
   (f) (Cf. Example 3.25.) Let $a_1=0$, $a_2=1$, and $a_n=\frac{2}{3}a_{n-2}+\frac{1}{3}a_{n-1}$ for $n>2$.
   (g) Let $a_1=1$ and $a_{n+1}=\frac{a_n}{2}+\frac{1}{a_n}$.
4. We can use Cauchy sequences to obtain roots of polynomials. E.g. using a graphing calculator, we see that x^3-4x+2 has exactly one root, call it a, in the interval $[0,1]$.
   (i) Show that the root a satisfies $a=\frac{1}{4}(a^3+2)$.
   (ii) Define a sequence $\{a_n\}$ recursively by $a_{n+1}=\frac{1}{4}(a_n^3+2)$ with $a_1=0$.
   (iii) Prove that $\{a_n\}$ is contractive and converges to a.
5. Here are some Cauchy limit theorems. Let $\{a_n\}$ be a sequence in $\mathbb{R}^m$.
   (a) Prove that $\{a_n\}$ is Cauchy if and only if for every $\varepsilon>0$ there is a number N such that for all $n>N$ and $k\geq 1$, $|a_{n+k}-a_n|<\varepsilon$.
   (b) Given any sequence $\{b_n\}$ of natural numbers, we call the sequence $\{d_n\}$, where $d_n=a_{n+b_n}-a_n$, a **difference sequence**. Prove that $\{a_n\}$ is Cauchy if and only if *every* difference sequence converges to zero (that is, is a null sequence). Suggestion: To prove the "if" part, instead prove the contrapositive: If $\{a_n\}$ is not Cauchy, then there is a difference sequence that does not converge to zero.

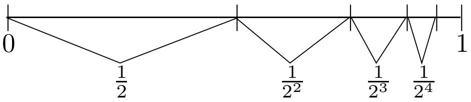


FIGURE 3.3. A stick of length one is halved infinitely many times.

6. (**Continued fractions** — see Chapter 8 for more on this amazing topic!) In this problem we investigate the **continued fraction**

$$\cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}}}.$$

We interpret this infinite fraction as the limit of the fractions

$$a_1 = \frac{1}{2}, \quad a_2 = \frac{1}{2 + \frac{1}{2}}, \quad a_3 = \frac{1}{2 + \frac{1}{2 + \frac{1}{2}}}, \ldots .$$

In other words, this sequence is defined by $a_1 = \frac{1}{2}$ and $a_{n+1} = \frac{1}{2+a_n}$ for $n \geq 1$. Prove that $\{a_n\}$ converges and find its limit. Here's a related example: Prove that

$$\boxed{\Phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}}} \tag{3.36}$$

in the sense that the right-hand continued fraction converges with value Φ. More precisely, prove that $\Phi - 1 = \lim \phi_n$ where $\{\phi_n\}$ is the sequence defined by $\phi_1 := 1$ and $\phi_{n+1} = 1/(1 + \phi_n)$ for $n \in \mathbb{N}$.

3.5. Baby infinite series

Imagine taking a stick of length 1 foot and cutting it in half, getting two sticks of length 1/2. We then take one of the halves and cut this piece in half, getting two sticks of length $1/4 = 1/2^2$. We now take one of these fourths and cut it in half, getting two sticks of length $1/8 = 1/2^3$. We continue this process indefinitely. Then, see Figure 3.3, the sum of all the lengths of all the sticks is 1:

$$1 = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} + \cdots .$$

In this section we introduce the theory of infinite series, which rigorously defines the right-hand sum.[4]

[4] *If you disregard the very simplest cases, there is in all of mathematics not a single infinite series whose sum has been rigorously determined. In other words, the most important parts of mathematics stand without a foundation. Niels H. Abel (1802–1829)* [**210**]. (Of course, nowadays series are rigorously determined — this is the point of this section!)

3.5.1. Basic results on infinite series. Given a sequence $\{a_n\}_{n=1}^{\infty}$ of complex numbers, we want to attach a meaning to $\sum_{n=1}^{\infty} a_n$, mostly written $\sum a_n$ for simplicity. To do so, we define the **n-th partial sum**, s_n, of the series to be

$$s_n := \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

Of course, here there are only finitely many numbers being summed, so the right-hand side has a clear definition. If the sequence $\{s_n\}$ of partial sums converges, then we say that the infinite series $\sum a_n$ **converges** and we define

$$\sum a_n = \sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots := \lim s_n.$$

If the sequence of partial sums does not converge to a complex number, then we say that the series **diverges**. Since $\mathbb{R} \subseteq \mathbb{C}$, restricting to real sequences $\{a_n\}$, we already have built in to the above definition the convergence of a series of real numbers. Just as a sequence can be indexed so its starting value is a_0 or a_{-7} or a_{1234}, etc., we can also consider series starting with indices other than 1:

$$\sum_{n=0}^{\infty} a_n, \qquad \sum_{n=-7}^{\infty} a_n, \qquad \sum_{n=1234}^{\infty} a_n, \qquad \text{etc.}$$

For convenience, in all our proofs we shall most of the time work with series starting at $n = 1$, although all the results we shall discuss work for series starting with any index.

Example 3.28. Consider the series

$$\sum_{n=0}^{\infty} (-1)^n = 1 - 1 + 1 - 1 + - \cdots .$$

Observe that $s_1 = 1$, $s_2 = 1 - 1 = 0$, $s_3 = 1 - 1 + 1 = 1$, and in general, $s_n = 1$ if n is odd and $s_n = 0$ if n is even. Since $\{s_n\}$ diverges, the series $\sum_{n=0}^{\infty}(-1)^n$ diverges.

There are two very simple tests that will help to determine the convergence or divergence of a series. The first test might also be called the "fundamental test" because it is the first thing that one should always test when given a series.

Theorem 3.19 (**n-th term test**). *If $\sum a_n$ converges, then $\lim a_n = 0$. Stated another way, if $\lim a_n \not\to 0$, then $\sum a_n$ diverges.*

Proof. Let $s = \sum a_n$ and s_n denote the n-th partial sum of the series. Observe that

$$s_n - s_{n-1} = a_1 + a_2 + \cdots + a_{n-1} + a_n - (a_1 + a_2 + \cdots + a_{n-1}) = a_n.$$

By definition of convergence of $\sum a_n$, we have $s_n \to s$. Therefore, $s_{n-1} \to s$ as well, hence $a_n = s_n - s_{n-1} \to s - s = 0$. □

Example 3.29. So, for example, the series $\sum_{n=0}^{\infty}(-1)^n = 1 - 1 + 1 - 1 + - \cdots$ and $\sum_{n=1}^{\infty} n = 1 + 2 + 3 + \cdots$ cannot converge, since their n-th terms do not tend to zero.

The converse of the n-th term test is false; that is, even though $\lim a_n = 0$, it may not follow that $\sum a_n$ exists.[5] For example, the ...

Example 3.30. (**Harmonic series diverges, Proof I**) Consider

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots .$$

This series is called the **harmonic series**; see [**125**] for "what's harmonic about the harmonic series". To see that the harmonic series does not converge, observe that

$$\begin{aligned} s_{2n} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6}\right) + \cdots + \left(\frac{1}{2n-1} + \frac{1}{2n}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{6} + \frac{1}{6}\right) + \cdots + \left(\frac{1}{2n} + \frac{1}{2n}\right) \\ &= 1 + \frac{1}{2} + \left(\frac{1}{2}\right) + \left(\frac{1}{3}\right) + \cdots + \left(\frac{1}{n}\right) = \frac{1}{2} + s_n. \end{aligned}$$

Thus, $s_{2n} > 1/2 + s_n$. Now if the harmonic series did converge, say to some real number s, that is, $s_n \to s$, then we would also have $s_{2n} \to s$. However, according to the inequality above, this would imply that $s \geq 1/2 + s$, which is an impossibility. Therefore, the harmonic series does not converge. See Problem 5 for more proofs.

Using the inequality $s_{2n} > 1/2 + s_n$, one can show (and we encourage you to do it!) that the partial sums of the harmonic series are unbounded. Then one can deduce that the harmonic series must diverge by the following very useful test.

Theorem 3.20 (**Nonnegative series test**). *A series $\sum a_n$ of nonnegative real numbers converges if and only if the sequence $\{s_n\}$ of partial sums is bounded, in which case, $s_n \leq s$ for all n where $s = \sum a_n := \lim s_n$.*

Proof. Since $a_n \geq 0$ for all n, we have

$$s_n = a_1 + a_2 + \cdots + a_n \leq a_1 + a_2 + \cdots + a_n + a_{n+1} = s_{n+1},$$

so the sequence of partial sums $\{s_n\}$ is nondecreasing: $s_1 \leq s_2 \leq \cdots \leq s_n \leq \cdots$. By the monotone criterion for sequences, the sequence of partial sums converges if and only if it is bounded. To see that $s_n \leq s := \sum_{m=1}^{\infty} a_m$ for all n, fix $n \in \mathbb{N}$ and note that $s_n \leq s_k$ for all $k \geq n$ because the partial sums are nondecreasing. Taking $k \to \infty$ and using that limits preserve inequalities gives $s_n \leq \lim s_k = s$. □

Example 3.31. Consider the following series:

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

To analyze this series, we use the "method of partial fractions" and note that $\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$. Thus, the adjacent terms in s_n cancel (except for the first and

[5] *The sum of an infinite series whose final term vanishes perhaps is infinite, perhaps finite. Jacob Bernoulli (1654–1705) Ars conjectandi.*

the last):

$$\begin{aligned}(3.37)\quad s_n &= \frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\cdots+\frac{1}{n(n+1)}\\ &= \left(\frac{1}{1}-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n}-\frac{1}{n+1}\right)=1-\frac{1}{n+1}\le 1.\end{aligned}$$

Hence, the sequence $\{s_n\}$ is bounded above by 1, so our series converges. Moreover, we also see that $s_n = 1 - 1/(n+1) \to 1$, and therefore

$$\sum_{n=1}^{\infty}\frac{1}{n(n+1)} = 1.$$

Example 3.32. Now if the sum of the reciprocals of the natural numbers diverges, what about the sum of the reciprocals of the squares (called the 2-series):

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = 1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots .$$

To investigate the convergence of this 2-series, using (3.37) we note that

$$\begin{aligned}s_n &= 1+\frac{1}{2\cdot 2}+\frac{1}{3\cdot 3}+\frac{1}{4\cdot 4}+\cdots+\frac{1}{n\cdot n}\\ &\le 1+\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots+\frac{1}{(n-1)\cdot n}\le 1+1=2.\end{aligned}$$

Since the partial sums of the 2-series are bounded, the 2-series converges. Now what is the value of this series? This question was answered by Leonhard Euler (1707–1783) in 1734. We shall rigourously prove, in 9 different ways in this book that the value of the 2-series is $\pi^2/6$ starting in Section 6.11! (Now what does π have to do with reciprocals of squares of natural numbers???)

3.5.2. Some properties of series. It is important to understand that the convergence or divergence of a series only depends on the "tails" of the series.

Theorem 3.21 (**Tails theorem for series**). *A series $\sum a_n$ converges if and only if there is an index m such that $\sum_{n=m}^{\infty} a_n$ converges.*

Proof. Let s_n denote the n-th partial sum of $\sum a_n$ and t_n that of any "m-tail" $\sum_{n=m}^{\infty} a_n$. Then

$$t_n = \sum_{k=m}^{n} a_k = s_n - a,$$

where a is the number $a = \sum_{k=1}^{m-1} a_k$. It follows that $\{s_n\}$ converges if and only if $\{t_n\}$ converges and our theorem is proved. □

An important type of series we'll run into often are geometric series. Given a complex number a, the series $\sum a^n$ is called a **geometric series**. The following theorem characterizes those geometric series that converge.

Theorem 3.22 (**Geometric series theorem**). *For any nonzero complex number a and $k \in \mathbb{Z}$, the geometric series $\sum_{n=k}^{\infty} a^n$ converges if and only if $|a| < 1$, in which case*

$$\sum_{n=k}^{\infty} a^n = a^k + a^{k+1} + a^{k+2} + a^{k+3} + \cdots = \frac{a^k}{1-a}.$$

PROOF. If $|a| \geq 1$ and $n \geq 0$, then $|a|^n \geq 1$, so the terms of the geometric series do not tend to zero, and therefore the geometric series cannot converge by the n-th term test. Thus, we may henceforth assume that $|a| < 1$. By the formula for a geometric progression (see (2.3) in Section 2.2), we have

$$s_n = a^k + a^{k+1} + \cdots + a^{k+n} = a^k \Big(1 + a + \cdots + a^n \Big) = a^k \, \frac{1 - a^{n+1}}{1 - a}.$$

Since $|a| < 1$, we know that $\lim a^{n+1} = 0$, so $s_n \to a^k/(1-a)$. This shows that the geometric series converges with sum equal to $a^k/(1-a)$. □

Of course, if $a = 0$, then the geometric series $a^k + a^{k+1} + a^{k+2} + \cdots$ is not defined if $k \leq 0$ and equals zero if $k \geq 1$.

Example 3.33. If we put $a = 1/2 < 1$ in the geometric series theorem, then we have

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots = \frac{1/2}{1 - 1/2} = 1,$$

just as hypothesized in the introduction to this section!

Finally, we state the following theorem on linear combinations of series.

THEOREM 3.23 (**Arithmetic properties of series**). *If $\sum a_n$ and $\sum b_n$ converge, then given any complex numbers, c, d, the series $\sum (c\,a_n + d\,b_n)$ converges, and*

$$\sum (c\,a_n + d\,b_n) = c \sum a_n + d \sum b_n.$$

Moreover, we can group the terms in the series

$$a_1 + a_2 + a_3 + a_4 + a_5 + \cdots$$

inside parentheses in any way we wish as long as we do not change the ordering of the terms and the resulting series still converges with sum $\sum a_n$; in other words, the associative law holds for convergent infinite series.

PROOF. The n-th partial sum of $\sum (c\,a_n + d\,b_n)$ is

$$\sum_{k=1}^{n} (c\,a_k + d\,b_k) = c \sum_{k=1}^{n} a_k + d \sum_{k=1}^{n} b_k = c\,s_n + d\,t_n,$$

where s_n and t_n are the n-th partial sums of $\sum a_n$ and $\sum b_n$, respectively. Since $s_n \to \sum a_n$ and $t_n \to \sum b_n$, the first statement our theorem follows.

Let $1 = \nu_1 < \nu_2 < \nu_3 < \cdots$ be any strictly increasing sequence of integers. We must show that

$$(a_1 + a_2 + \cdots + a_{\nu_2 - 1}) + (a_{\nu_2} + a_{\nu_2 + 1} + \cdots + a_{\nu_3 - 1}) +$$
$$(a_{\nu_4} + a_{\nu_4} + \cdots + a_{\nu_5 - 1}) + \cdots =: \sum_{n=1}^{\infty} (a_{\nu_n} + \cdots + a_{\nu_{n+1} - 1})$$

converges with sum $\sum_{n=1}^{\infty} a_n$. To see this, observe that if $\{s_n\}$ denotes the partial sums of $\sum_{n=1}^{\infty} a_n$ and if $\{S_n\}$ denotes the partial sums of the right-hand series, then

$$S_n = (a_1 + a_2 + \cdots + a_{\nu_2 - 1}) + \cdots + (a_{\nu_n} + a_{\nu_4} + \cdots + a_{\nu_{n+1} - 1}) = s_{\nu_{n+1} - 1},$$

since the associative law holds for finite sums. Therefore, $\{S_n\}$ is just a subsequence of $\{s_n\}$, and hence has the same limit: $\sum_{n=1}^{\infty} (a_{\nu_n} + \cdots + a_{\nu_{n+1} - 1}) = \lim S_n = \lim s_n = \sum a_n$. This completes our proof. □

In Section 6.6, we'll see that the commutative law doesn't hold! It is worth remembering that the associative law does not work in reverse.

Example 3.34. For instance, the series

$$0 = 0 + 0 + 0 + 0 + \cdots = (1 - 1) + (1 - 1) + (1 - 1) + \cdots$$

certainly converges, but we cannot omit the parentheses and conclude that $1 - 1 + 1 - 1 + 1 - 1 + \cdots$ converges, which we already showed does not.

3.5.3. Telescoping series. As seen in Example 3.31, the value of the series $\sum 1/n(n+1)$ was very easy to find because in writing out its partial sums, we saw that the sum "telescoped" to give a simple expression. In general, it is very difficult to find the value of a convergent series, but for telescoping series, the sums are quite straightforward to find.

Theorem 3.24 (**Telescoping series theorem**). *Let $\{x_n\}$ be a sequence of complex numbers and let $\sum_{n=0}^{\infty} a_n$ be the series with n-th term $a_n := x_n - x_{n+1}$. Then* $\lim x_n$ *exists if and only if $\sum a_n$ converges, in which case*

$$\sum_{n=0}^{\infty} a_n = x_0 - \lim x_n.$$

Proof. Observe that adjacent terms of the following partial sum cancel:

$$s_n = (x_0 - x_1) + (x_1 - x_2) + \cdots + (x_{n-1} - x_n) + (x_n - x_{n+1}) = x_0 - x_{n+1}.$$

If $x := \lim x_n$ exists, we have $x = \lim x_{n+1}$ as well, and therefore $\sum a_n := \lim s_n$ exists with sum $x_0 - x$. Conversely, if $s = \lim s_n$ exists, then $s = \lim s_{n-1}$ as well, and since $x_n = x_0 - s_{n-1}$, it follows that $\lim x_n$ exists. □

Example 3.35. Let a be any nonzero complex number not equal to a negative integer. Then we claim that

$$\sum_{n=0}^{\infty} \frac{1}{(n+a)(n+a+1)} = \frac{1}{a}.$$

Indeed, in this case, we can use the "method of partial fractions" to write

$$a_n = \frac{1}{(n+a)(n+a+1)} = \frac{1}{(n+a)} - \frac{1}{n+a+1} = x_n - x_{n+1}, \text{ where } x_n = \frac{1}{(n+a)}.$$

Since $\lim x_n = 0$, applying the telescoping series theorem gives

$$\sum_{n=0}^{\infty} \frac{1}{(n+a)(n+a+1)} = x_0 = \frac{1}{(0+a)} = \frac{1}{a},$$

just as we stated.

Example 3.36. More generally, given any natural number k we have

$$\boxed{\sum_{n=0}^{\infty} \frac{1}{(n+a)(n+a+1)\cdots(n+a+k)} = \frac{1}{k}\,\frac{1}{a(a+1)\cdots(a+k-1)}.} \tag{3.38}$$

Indeed, in this general case, we have (again using partial fractions)

$$\frac{1}{(n+a)(n+a+1)\cdots(n+a+k)} = \underbrace{\frac{1}{k(n+a)\cdots(n+a+k-1)}}_{x_n} - \underbrace{\frac{1}{k(n+a+1)\cdots(n+a+k)}}_{x_{n+1}}.$$

Since $\lim x_n = 0$ (why?), applying the telescoping series theorem gives

$$\sum_{n=0}^{\infty} \frac{1}{(n+a)(n+a+1)\cdots(n+a+k)} = x_0 = \frac{1}{k}\,\frac{1}{a(a+1)\cdots(a+k-1)},$$

just as we stated. For example, with $a = 1/2$ and $k = 2$ in (3.38), we obtain (after a little algebra)

$$\boxed{\frac{1}{1\cdot 3\cdot 5} + \frac{1}{3\cdot 5\cdot 7} + \frac{1}{5\cdot 7\cdot 9} + \cdots = \frac{1}{12}}$$

and with $a = 1/3$ and $k = 2$ in (3.38), we obtain another beautiful sum:

$$\boxed{\frac{1}{1\cdot 4\cdot 7} + \frac{1}{4\cdot 7\cdot 10} + \frac{1}{7\cdot 10\cdot 13} + \cdots = \frac{1}{24}.}$$

More examples of telescoping series can be found in the article [**184**]. In summary, the telescoping series theorem is useful in quickly finding sums to certain series. Moreover, this theorem allows us to construct series with any specified sum.

Corollary 3.25. *Let s be any complex number and let $\{x_n\}_{n=0}^{\infty}$ be a null sequence (that is, $\lim x_n = 0$) such that $x_0 = s$. Then setting $a_n := x_n - x_{n+1}$, the series $\sum_{n=0}^{\infty} a_n$ converges to s.*

Proof. By the telescoping series theorem, $\sum a_n = x_0 - \lim x_n = s - 0 = s$. □

Example 3.37. For example, let $s = 1$. Then $x_n = 1/2^n$ defines a null sequence with $1/2^0 = 1$, so $\sum_{n=0}^{\infty} a_n = 1$ with

$$a_n = x_n - x_{n+1} = \frac{1}{2^n} - \frac{1}{2^{n+1}} = \frac{1}{2^{n+1}}.$$

Hence,

$$1 = \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} = \sum_{n=1}^{\infty} \frac{1}{2^n},$$

a fact that we already knew from our work with geometric series.

Example 3.38. Also, $x_n = 1/(n+1)$ defines a null sequence with $x_0 = 1$. In this case,

$$a_n = x_n - x_{n+1} = \frac{1}{n+1} - \frac{1}{n+2} = \frac{1}{(n+1)\,(n+2)},$$

so

$$1 = \sum_{n=0}^{\infty} \frac{1}{(n+1)\,(n+2)} = \sum_{n=1}^{\infty} \frac{1}{n\,(n+1)},$$

another fact we already knew!

What fancy formulas for 1 do you get when you put $x_n = 1/(n+1)^2$ and $x_n = 1/\sqrt{n+1}$?

EXERCISES 3.5.

1. Determine the convergence of the following series. If the series converges, find the sum.

$$(a)\ \sum_{n=1}^{\infty}\left(1+\frac{1}{n}\right)^n \quad , \quad (b)\ \sum_{n=1}^{\infty}\left(\frac{i}{2}\right)^n \quad , \quad (c)\ \sum_{n=1}^{\infty}\frac{1}{n^{1/n}}.$$

2. Let $\{a_n\}$ be a sequence of complex numbers.
   (a) Assume that $\sum a_n$ converges. Prove that the sum of the even terms $\sum_{n=1}^{\infty} a_{2n}$ converges if and only if the sum of the odd terms $\sum_{n=1}^{\infty} a_{2n-1}$ converges, in which case, $\sum a_n = \sum a_{2n} + \sum a_{2n-1}$.
   (b) Let $\sum c_n$ be a series obtained from $\sum a_n$ by modifying at most finitely many terms. Show that $\sum a_n$ converges if and only if $\sum c_n$ converges.
   (c) Assume that $\lim a_n = 0$. Fix $\alpha, \beta \in \mathbb{C}$ with $\alpha + \beta \neq 0$. Prove that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty}(\alpha a_n + \beta a_{n+1})$ converges.
3. Prove that

$$\sum_{n=1}^{\infty}\frac{n}{2^n} = \frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots = 2.$$

Suggestion: Problem 3d in Exercises 2.2 might help.

4. Let a be a complex number. Using the telescoping series theorem, show that

$$\boxed{\frac{a}{1-a^2} + \frac{a^2}{1-a^4} + \frac{a^4}{1-a^8} + \frac{a^8}{1-a^{16}} + \cdots = \begin{cases} \frac{a}{1-a} & |a|<1 \\ \frac{1}{1-a} & |a|>1. \end{cases}}$$

Suggestion: Using the identity $\frac{x}{1-x^2} = \frac{1}{1-x} - \frac{1}{1-x^2}$ for any $x \neq \pm 1$ (you should prove this identity!), write $\frac{a^{2^n}}{1-a^{2^{n+1}}}$ as the difference $x_n - x_{n+1}$ where $x_n = 1 - a^{2^n}$.

5. ($\sum_{n=1}^{\infty}\frac{1}{n}$ **diverges, Proofs II–IV**) For more proofs, see [**115**]. Let $s_n = \sum_{k=1}^{n}\frac{1}{k}$.
   (a) Using that $1 + \frac{1}{n} < e^{1/n}$ for all $n \in \mathbb{N}$, which is from (3.28), show that

$$\left(1+\frac{1}{1}\right)\left(1+\frac{1}{2}\right)\left(1+\frac{1}{3}\right)\cdots\left(1+\frac{1}{n}\right) \le e^{s_n} \ , \quad \text{for all } \ n \in \mathbb{N}.$$

   Show that the left-hand side equals $n+1$ and conclude that $\{s_n\}$ cannot converge.
   (b) Show that for any $k \in \mathbb{N}$ with $k \geq 3$, we have

$$\frac{1}{k-1} + \frac{1}{k} + \frac{1}{k+1} \geq \frac{3}{k}.$$

   Using this inequality, prove that for any $n \in \mathbb{N}$, $s_{3n+1} \geq 1 + s_n$ by grouping the terms of s_{3n+1} into threes (starting from $\frac{1}{2}$). Now show that $\{s_n\}$ cannot converge.
   (c) For any $k \in \mathbb{N}$ with $k \geq 2$, prove the inequality

$$\frac{1}{(k-1)!+1} + \frac{1}{(k-1)!+2} + \cdots + \frac{1}{k!} \geq 1 - \frac{1}{k}.$$

   Writing $s_{n!}$ into groups of the form given on the left-hand side of this inequality, prove that $s_{n!} \geq 1 + n - s_n$. Conclude that $\{s_n\}$ cannot converge.
6. We shall prove that $\sum_{n=1}^{\infty} nz^{n-1}$ converges if and only if $|z|<1$, in which case

$$\frac{1}{(1-z)^2} = \sum_{n=1}^{\infty} nz^{n-1}. \tag{3.39}$$

   (i) Prove that if $\sum_{n=1}^{\infty} nz^{n-1}$ converges, then $|z|<1$.
   (ii) Prove that $(1-z)\sum_{k=1}^{n} kz^{k-1}$ can be written as

$$(1-z)\sum_{k=1}^{n} kz^{k-1} = \frac{1-z^n}{1-z} - nz^n.$$

(iii) Now prove that if $|z| < 1$, then $1/(1-z)^2 = \sum_{n=1}^\infty nz^{n-1}$. Solve Problem 3 using (3.39). Suggestion: Problem 4 in Exercises 3.1 might be helpful.

(iv) Can you prove that $\frac{2}{(1-z)^3} = \sum_{n=2}^\infty n(n-1)z^{n-2}$ for $|z| < 1$ using a similar technique? (Do this problem if you are feeling extra confident!)

7. In Problem 9 of Exercises 2.2 we studied the **Fibonacci sequence**, $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. Using the telescoping theorem prove that

$$\sum_{n=2}^\infty \frac{1}{F_{n-1}F_{n+1}} = 1 \quad , \quad \sum_{n=2}^\infty \frac{F_n}{F_{n-1}F_{n+1}} = 2.$$

8. Here is a generalization of the telescoping series theorem.

(i) Let $x_n \to x$ and let $k \in \mathbb{N}$. Prove that $\sum_{n=0}^\infty a_n$ with $a_n = x_n - x_{n+k}$ converges and has the sum

$$\sum_{n=0}^\infty a_n = x_0 + x_1 + \cdots + x_{k-1} - k\,x.$$

Using this formula, find the sums

$$\sum_{n=0}^\infty \frac{1}{(2n+1)(2n+7)} \qquad \text{and} \qquad \sum_{n=0}^\infty \frac{1}{(3n+1)(3n+7)}.$$

(ii) Let a be any nonzero complex number not equal to a negative integer and let $k \in \mathbb{N}$. Using (i), prove that

$$\boxed{\sum_{n=0}^\infty \frac{1}{(n+a)(n+a+k)} = \frac{1}{k}\left[\frac{1}{a} + \frac{1}{a+1} + \cdots + \frac{1}{a+k-1}\right].}$$

With $a = 1$ and $k = 2$, derive a beautiful expression for $3/4$.

(iii) Here's a fascinating result: Given another natural number m, prove that

$$\boxed{\begin{aligned}&\sum_{n=0}^\infty \frac{1}{(n+a)(n+a+m)\cdots(n+a+km)}\\ &\qquad = \frac{1}{km}\sum_{n=0}^{m-1} \frac{1}{(n+a)(n+a+m)\cdots(n+a+(k-1)m)}.\end{aligned}}$$

Find a beautiful series when $a = 1$ and $k = m = 2$.

9. Let $x_n \to x$, and let $c_1, \ldots, c_k$ be $k \geq 2$ numbers such that $c_1 + \cdots + c_k = 0$.

(i) Prove that the series $\sum_{n=0}^\infty a_n$ with $a_n = c_1\,x_{n+1} + c_2\,x_{n+2} + \cdots + c_k\,x_{n+k}$ converges and has the sum

$$\sum_{n=0}^\infty a_n = c_1\,x_1 + (c_1 + c_2)\,x_2 + \cdots + (c_1 + c_2 + \cdots + c_{k-1})\,x_{k-1} + (c_2 + 2\,c_3 + 3c_4 + \cdots + (k-1)\,c_k)\,x.$$

(ii) Using (i), find the sum of

$$\frac{5}{5\cdot 7\cdot 9} + \frac{11}{7\cdot 9\cdot 11} + \frac{17}{9\cdot 11\cdot 13} + \cdots + \frac{6n+5}{(2n+5)(2n+7)(2n+9)} + \cdots.$$

10. Let a be any complex number not equal to $0, -1, -1/2, -1/3, \ldots$. Prove that

$$\sum_{n=1}^\infty \frac{n}{(a+1)(2a+1)\cdots(na+1)} = \frac{1}{a}.$$

3.6. Absolute convergence and a potpourri of convergence tests

We now give some important tests that guarantee when certain series converge.

3.6.1. Various tests for convergence. The first test is the series version of Cauchy's criterion for sequences.

THEOREM 3.26 (**Cauchy's criterion for series**). *The series $\sum a_n$ converges if and only if given any $\varepsilon > 0$ there is an N such that for all $n > m > N$, we have*

$$\left| \sum_{k=m+1}^{n} a_k \right| = |a_{m+1} + a_{m+2} + \cdots + a_n| < \varepsilon,$$

in which case, for any $m > N$,

$$\left| \sum_{k=m+1}^{\infty} a_k \right| < \varepsilon.$$

In particular, for a convergent series $\sum a_n$, we have

$$\lim_{m\to\infty} \sum_{k=m+1}^{\infty} a_k = 0.$$

PROOF. Let s_n denote the n-th partial sum of $\sum a_n$. Then to say that the series $\sum a_n$ converges means that the sequence $\{s_n\}$ converges. Cauchy's criterion for sequences states that $\{s_n\}$ converges if and only if given any $\varepsilon > 0$ there is an N such that for all $n, m > N$, we have $|s_n - s_m| < \varepsilon$. As n and m are symmetric in this criterion we may assume that $n > m > N$. Since

$$s_n - s_m = \sum_{k=1}^{n} a_k - \sum_{k=1}^{m} a_k = \sum_{k=m+1}^{n} a_k,$$

this Cauchy criterion is equivalent to $|\sum_{k=m+1}^{n} a_k| < \varepsilon$ for all $n > m \geq N$. Taking $n \to \infty$ shows that $|\sum_{k=m+1}^{\infty} a_k| < \varepsilon$. □

Here is another test, which is the most useful of the ones we've looked at.

THEOREM 3.27 (**Comparison test**). *Let $\{a_n\}$ and $\{b_n\}$ be real sequences and suppose that for n sufficiently large, say for all $n \geq k$ for some $k \in \mathbb{N}$, we have*

$$0 \leq a_n \leq b_n.$$

If $\sum b_n$ converges, then $\sum a_n$ converges. Equivalently, if $\sum a_n$ diverges, then $\sum b_n$ diverges. In the case of convergence, $\sum_{n=k}^{\infty} a_n \leq \sum_{n=k}^{\infty} b_n$.

PROOF. By the tails theorem for series (Theorem 3.21), the series $\sum a_n$ and $\sum b_n$ converge if and only if the series $\sum_{n=k}^{\infty} a_n$ and $\sum_{n=k}^{\infty} b_n$ converge. By working with these series instead of the original ones, we may assume that $0 \leq a_n \leq b_n$ holds for every n. In this case, if s_n denotes the n-partial sum for $\sum a_n$ and t_n, the n-th partial sum for $\sum b_n$, then $0 \leq a_n \leq b_n$ (for every n) implies that for every n, we have

$$0 \leq s_n \leq t_n.$$

Assume that $\sum b_n$ converges. Then by the nonnegative series test (Theorem 3.20), $t_n \leq t := \sum b_n$. Hence, $0 \leq s_n \leq t$ for all n; that is, the partial sums of $\sum a_n$ are bounded. Again by the nonnegative series test, it follows that $\sum a_n$ converges and taking $n \to \infty$ in $0 \leq s_n \leq t$ shows that $0 \leq \sum a_n \leq \sum b_n$. □

Example 3.39. For example, the **p-series**, where p is a rational number,

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots ,$$

converges for $p \geq 2$ and diverges for $p \leq 1$. To see this, note that if $p < 1$, then

$$\frac{1}{n} \leq \frac{1}{n^p},$$

because $1 - p > 0$, so $1 = 1^{1-p} \leq n^{1-p}$ by the power rules theorem (Theorem 2.33) and $1 \leq n^{1-p}$ is equivalent to the above inequality. Since the harmonic series diverges, by the comparison test, so does the p-series for $p < 1$. If $p > 2$, then by a similar argument, we have

$$\frac{1}{n^p} \leq \frac{1}{n^2}.$$

In the last section, we showed that the 2-series $\sum 1/n^2$ converges, so by the comparison test, the p-series for $p > 2$ converges. Now what about for $1 < p < 2$? To answer this question we shall appeal to Cauchy's condensation test below.

3.6.2. Cauchy condensation test. The following test is usually not found in elementary calculus textbooks, but it's very useful.

Theorem 3.28 (**Cauchy condensation test**). *If $\{a_n\}$ is a nonincreasing sequence of nonnegative real numbers, then the infinite series $\sum a_n$ converges if and only if*

$$\sum_{n=0}^{\infty} 2^n a_{2^n} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

converges.

Proof. The proof of this theorem is just like the proof of the p-test! Let the partial sums of $\sum a_n$ be denoted by s_n and those of $\sum_{n=0}^{\infty} 2^n a_{2^n}$ by t_n. Then by the nonnegative series test (Theorem 3.20), $\sum a_n$ converges if and only if $\{s_n\}$ is bounded and $\sum 2^n a_{2^n}$ converges if and only if $\{t_n\}$ is bounded. Therefore, we just have to prove that $\{s_n\}$ is bounded if and only if $\{t_n\}$ is bounded.

Consider the "if" part: Assume that $\{t_n\}$ is bounded; we shall prove that $\{s_n\}$ is bounded. To prove this, we note that $s_n \leq s_{2^n - 1}$ and we can write (cf. the above computation with the p-series)

$$s_n \leq s_{2^n-1} = a_1 + (a_2 + a_3) + (a_4 + a_5 + a_6 + a_7) + \cdots + (a_{2^{n-1}} + \cdots + a_{2^n - 1}),$$

where in the k-th parenthesis, we group those terms of the series with index running from 2^k to $2^{k+1} - 1$. Since the a_n's are nonincreasing (that is, $a_n \geq a_{n+1}$ for all n), replacing each number in a parenthesis by the first term in the parenthesis we can only increase the value of the sum, so

$$\begin{aligned} s_n &\leq a_1 + (a_2 + a_2) + (a_4 + a_4 + a_4 + a_4) + \cdots + (a_{2^{n-1}} + \cdots + a_{2^{n-1}}) \\ &\leq a_1 + 2a_2 + 4a_4 + \cdots + 2^{n-1} a_{2^{n-1}} = t_{n-1}. \end{aligned}$$

Since $\{t_n\}$ is bounded, it follows that $\{s_n\}$ is bounded as well.

Now the "only if" part: Assume that $\{s_n\}$ is bounded; we shall prove that $\{t_n\}$ is bounded. To prove this, we try to estimate t_n using s_{2^n}. Observe that

$$\begin{aligned}
s_{2^n} &= a_1 + a_2 + (a_3 + a_4) + (a_5 + a_6 + a_7 + a_8) + \cdots + (a_{2^{n-1}+1} + \cdots + a_{2^n})\\
&\geq a_1 + a_2 + (a_4 + a_4) + (a_8 + a_8 + a_8 + a_8) + \cdots + (a_{2^n} + \cdots + a_{2^n})\\
&= a_1 + a_2 + 2a_4 + 4a_8 + \cdots + 2^{n-1} a_{2^n}\\
&= \frac{1}{2} a_1 + \frac{1}{2}\left(a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots + 2^n a_{2^n}\right)\\
&= \frac{1}{2} a_1 + \frac{1}{2} t_n.
\end{aligned}$$

It follows that $t_n \leq 2 s_{2^n}$ for all n. In particular, since $\{s_n\}$ is bounded, $\{t_n\}$ is bounded as well. This completes our proof. □

Example 3.40. For instance, consider the p-series (where $p \geq 0$ is rational):

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \cdots$$

With $a_n = \frac{1}{n^p}$, by Cauchy's condensation test, this series converges if and only if

$$\sum_{n=0}^{\infty} 2^n a_{2^n} = \sum_{n=1}^{\infty} \frac{2^n}{(2^n)^p} = \sum_{n=1}^{\infty} \left(\frac{1}{2^{p-1}}\right)^n$$

converges. This is a geometric series, so this series converges if and only if

$$\frac{1}{2^{p-1}} < 1 \qquad \Longleftrightarrow \qquad p > 1.$$

Summarizing, we get

$$\boxed{\textbf{p-test:} \quad \sum_{n=1}^{\infty} \frac{1}{n^p} \quad \begin{cases} \text{converges for } p > 1,\\ \text{diverges for } p \leq 1 \end{cases}}$$

Once we develop the theory of real exponents, the same p-test holds for p real. By the way, $\sum_{n=1}^{\infty} 1/n^p$ is also denoted by $\zeta(p)$, the **zeta function** at p:

$$\zeta(p) := \sum_{n=1}^{\infty} \frac{1}{n^p}.$$

We'll come across this function again in Section 4.6.

Cauchy's condensation test is especially useful when dealing with series involving logarithms; see the problems. Although we technically haven't introduced the logarithm function, we'll *thoroughly* develop this function in Section 4.6, so for now we'll assume you know properties of $\log x$ for $x > 0$. Actually, for the particular example below, we just need to know that $\log x^k = k \log x$ for all $k \in \mathbb{Z}$, $\log x > 0$ for $x > 1$, and $\log x$ is increasing with x.

Example 3.41. Consider the series

$$\sum_{n=2}^{\infty} \frac{1}{n \log n}.$$

At first glance, it may seem difficult to determine the convergence of this series, but Cauchy's condensation test gives the answer quickly:

$$\sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n \log 2^n} = \frac{1}{\log 2} \sum_{n=1}^{\infty} \frac{1}{n},$$

which diverges. (You should check that $1/(n \log n)$ is nonincreasing.) Therefore by Cauchy's condensation test, $\sum_{n=2}^{\infty} \frac{1}{n \log n}$ also diverges. [6]

3.6.3. Absolute convergence. A series $\sum a_n$ is said to be **absolutely convergent** if $\sum |a_n|$ converges. The following theorem implies that any absolutely convergent series is convergent in the usual sense.

Theorem 3.29 (**Absolute convergence**). *Let $\sum a_n$ be an infinite series.*

(1) If $\sum |a_n|$ converges, then $\sum a_n$ also converges, and

$$\left|\sum a_n\right| \leq \sum |a_n| \qquad \textit{(triangle inequality for series)}. \tag{3.40}$$

(2) Any linear combination of absolutely convergent series is absolutely convergent.

Proof. Suppose that $\sum |a_n|$ converges. We shall prove that $\sum a_n$ converges and (3.40) holds. To prove convergence, we use Cauchy's criterion, so let $\varepsilon > 0$. Since $\sum |a_n|$ converges, there is an N such that for all $n > m > N$, we have

$$\sum_{k=m+1}^{n} |a_k| < \varepsilon.$$

By the usual triangle inequality, for $n > m > N$, we have

$$\left|\sum_{k=m+1}^{n} a_k\right| \leq \sum_{k=m+1}^{n} |a_k| < \varepsilon.$$

Thus, by Cauchy's criterion for series, $\sum a_n$ converges. To prove (3.40), let s_n denote the n-th partial sum of $\sum a_n$. Then,

$$|s_n| = \left|\sum_{k=1}^{n} a_k\right| \leq \sum_{k=1}^{n} |a_k| \leq \sum_{k=1}^{\infty} |a_k|.$$

Since $|s_n| \to |\sum a_k|$, by the squeeze theorem it follows that $|\sum a_k| \leq \sum |a_k|$.

If $\sum |a_n|$ and $\sum |b_n|$ converge, then given any complex numbers c, d, as $|ca_n + db_n| \leq |c||a_n| + |d||b_n|$, by the comparison theorem, $\sum |ca_n + db_n|$ also converges. □

Example 3.42. Since the 2-series $\sum 1/n^2$ converges, each of the following series is absolutely convergent:

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}, \quad \sum_{n=1}^{\infty} \frac{i^n}{n^2}, \quad \sum_{n=1}^{\infty} \frac{(-i)^n}{n^2}.$$

It is possible to have a convergent series that is not absolutely convergent.

[6] This series is usually handled in elementary calculus courses using the technologically advanced "integral test," but Cauchy's condensation test gives one way to handle such series without knowing any calculus!

Example 3.43. Although the harmonic series $\sum 1/n$ diverges, the **alternating harmonic series**

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + - \cdots$$

converges. To see this, we use the Cauchy criterion. Given $n > m$, observe that

$$\left| \sum_{k=m+1}^{n} \frac{(-1)^{k-1}}{k} \right| = \left| \frac{(-1)^m}{m+1} + \frac{(-1)^{m+1}}{m+2} + \frac{(-1)^{m+2}}{m+3} + \cdots + \frac{(-1)^{n-1}}{n} \right|$$

$$= \left| \frac{1}{m+1} - \frac{1}{m+2} + \frac{1}{m+3} + \cdots + \frac{(-1)^{n-m-1}}{n} \right|. \tag{3.41}$$

Suppose that $n - m$ is even. Then the sum in the absolute values in (3.41) equals

$$\begin{aligned} &\frac{1}{m+1} - \frac{1}{m+2} + \frac{1}{m+3} - \frac{1}{m+4} + \cdots + \frac{1}{n-1} - \frac{1}{n} \\ &\quad = \left(\frac{1}{m+1} - \frac{1}{m+2} \right) + \left(\frac{1}{m+3} - \frac{1}{m+4} \right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n} \right) > 0, \end{aligned}$$

since all the terms in parentheses are positive. Thus, if $n - m$ is even, then we can drop the absolute values in (3.41) to get

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

since all the terms in parentheses are positive. Now suppose that $n - m$ is odd. Then the sum in the absolute values in (3.41) equals

$$\begin{aligned} &\frac{1}{m+1} - \frac{1}{m+2} + \frac{1}{m+3} - \frac{1}{m+4} + \cdots - \frac{1}{n-1} + \frac{1}{n} \\ &= \left(\frac{1}{m+1} - \frac{1}{m+2} \right) + \left(\frac{1}{m+3} - \frac{1}{m+4} \right) + \cdots + \left(\frac{1}{n-2} - \frac{1}{n-1} \right) + \frac{1}{n} > 0, \end{aligned}$$

since all the terms in parentheses are positive. So, if $n - m$ is odd, then just as before, we can drop the absolute values in (3.41) to get

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

since, once again, all the terms in parentheses are positive. In conclusion, regardless if $n - m$ is even or odd, we see that for any $n > m$, we have

$$\left| \sum_{k=m+1}^{n} \frac{(-1)^{k-1}}{k} \right| < \frac{1}{m+1}.$$

Since $1/(m+1) \to 0$ as $m \to \infty$, this inequality shows that the alternating harmonic series satisfies the conditions of Cauchy's criterion, and therefore converges. Another way to prove convergence is to use the "alternating series test," a subject we will study thoroughly in Section 6.1. Later on, in Section 4.6, we'll prove that $\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n}$ equals $\log 2$.

Example 3.44. Using the associative law in Theorem 3.23, we can also write

$$\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n} = \Big(1 - \frac{1}{2}\Big) + \Big(\frac{1}{3} - \frac{1}{4}\Big) + \Big(\frac{1}{5} - \frac{1}{6}\Big) = \frac{1}{1 \cdot 2} + \frac{1}{3 \cdot 4} + \frac{1}{5 \cdot 6} + \cdots .$$

Exercises 3.6.

1. For this problem, assume you know *all* the "well-known" high school properties of $\log x$ (e.g. $\log x^k = k \log x$, $\log(xy) = \log x + \log y$, etc.). Using the Cauchy condensation test, determine the convergence of the following series:
$$(a) \sum_{n=2}^\infty \frac{1}{n(\log n)^2} \quad , \quad (b) \sum_{n=2}^\infty \frac{1}{n(\log n)^p} \quad , \quad (c) \sum_{n=2}^\infty \frac{1}{n(\log n)\,(\log(\log n))}.$$
For *(b)*, state which p give convergent/diverent series.
2. Prove that
$$\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2 \cdot 3} - \frac{1}{4 \cdot 5} - \frac{1}{6 \cdot 7} - \cdots ,$$
$$\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n^2} = \frac{1}{1^2 \cdot 2^2}(1+2) + \frac{1}{3^2 \cdot 4^2}(3+4) + \frac{1}{5^2 \cdot 6^2}(5+6) + \cdots .$$
3. We consider various (unrelated) properties of real series $\sum a_n$ with $a_n \geq 0$ for all n.
   (a) Here is a nice generalization of the Cauchy condensation test: If the a_n's are nonincreasing, then given a natural number $b > 1$, prove that $\sum a_n$ converges or diverges with the series
   $$\sum_{n=0}^\infty b^n\, a_{b^n} = a_1 + b\, a_b + b^2\, a_{b^2} + b^3\, a_{b^3} + \cdots .$$
   Thus, the Cauchy condensation test is just this test with $b = 2$.
   (b) If $\sum a_n$ converges, prove that for any $k \in \mathbb{N}$, the series $\sum a_n^k$ also converges.
   (c) If $\sum a_n$ converges, give an example showing that $\sum \sqrt{a_n}$ may not converge. However, prove that the series $\sum \sqrt{a_n}/n$ does converges. Suggestion: Can you somehow use the AGMI with two terms?
   (d) If $a_n > 0$ for all n, prove that $\sum a_n$ converges if and only if for *any* series $\sum b_n$ of nonnegative real numbers, $\sum (a_n^{-1} + b_n)^{-1}$ converges.
   (e) If $\sum b_n$ is another series of nonnegative real numbers, prove that $\sum a_n$ and $\sum b_n$ converge if and only if $\sum \sqrt{a_n^2 + b_n^2}$ converges.
   (f) Prove that $\sum a_n$ converges if and only if $\sum \frac{a_n}{1+a_n}$ converges.
   (g) For each $n \in \mathbb{N}$, define $b_n = \sqrt{\sum_{k=n}^\infty a_k} = \sqrt{a_n + a_{n+1} + a_{n+2} + \cdots}$. Prove that if $a_n > 0$ for all n and $\sum a_n$ converges, then $\sum \frac{a_n}{b_n}$ converges. Suggestion: Show that $a_n = b_n^2 - b_{n+1}^2$ and using this fact show that $\frac{a_n}{b_n} \leq 2(b_n - b_{n+1})$ for all n.
4. We already know that if $\sum a_n$ (of complex numbers) converges, then $\lim a_n = 0$. When the a_n's form a nonincreasing sequence of nonnegative real numbers, then prove the following astonishing fact (called **Pringsheim's theorem**): If $\sum a_n$ converges, then $n\, a_n \to 0$. Use the Cauchy criterion for series somewhere in your proof. Suggestion: Let $\varepsilon > 0$ and choose N such that $n > m > N$ implies
$$a_{m+1} + a_{m+2} + \cdots + a_n < \frac{\varepsilon}{2}.$$
Take $n = 2m$ and then $n = 2m + 1$.
5. (**Limit comparison test**) Let $\{a_n\}$ and $\{b_n\}$ be nonzero complex sequences and suppose that the following limit exists: $L := \lim \left|\frac{a_n}{b_n}\right|$. Prove that

(i) If $L \neq 0$, then $\sum a_n$ is absolutely convergent if and only if $\sum b_n$ is absolutely convergent.

(ii) If $L = 0$ and $\sum b_n$ is absolutely convergent, then $\sum a_n$ is absolutely convergent.

6. Here's an alternative method to prove that the alternating harmonic series converges.
   (i) Let $\{b_n\}$ be a sequence in $\mathbb{R}^m$ and suppose that the even and odd subsequences $\{b_{2k}\}$ and $\{b_{2k-1}\}$ both converge and have the same limit L. Prove that the original sequence $\{b_n\}$ converges and has limit L.
   (ii) Show that the subsequences of even and odd partial sums of the alternating harmonic series both converge and have the same limit.
7. (**Ratio comparison test**) Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers and suppose that $\frac{a_{n+1}}{a_n} \leq \frac{b_{n+1}}{b_n}$ for all n. If $\sum b_n$ converges, prove that $\sum a_n$ also converges. (Equivalently, if $\sum a_n$ diverges, then $\sum b_n$ also diverges.) Suggestion: Consider the telescoping product
$$a_n = \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdot \frac{a_{n-2}}{a_{n-3}} \cdots \frac{a_2}{a_1} \cdot a_1.$$
8. (Cf. [**104**], [**113**], [**235**]) We already know that the harmonic series $\sum 1/n$ diverges. It turns out that omitting certain numbers from this sum makes the sum converge. Fix a natural number $b \geq 2$. Recall (see Section 2.5) that we can write any natural number n uniquely as $n = a_k a_{k-1} \cdots a_0$, where $0 \leq a_j \leq b-1$, $j = 0, \ldots, k$, are called digits, and where the notation $a_k \cdots a_0$ means that
$$a = a_k\, b^k + a_{k-1}\, b^{k-1} + \cdots + a_1\, b + a_0.$$
Prove that the following sum converges:
$$\sum_{n \text{ has no 0 digit}} \frac{1}{n}.$$
Suggestion: Let c_k be the sum over all numbers of the form $\frac{1}{n}$ where $n = a_k a_{k-1} \cdots a_0$ with none of a_j's zero. Show that there at most $(b-1)^{k+1}$ such n's and that $n \geq b^k$ and use these facts to show that $c_k \leq \frac{(b-1)^{k+1}}{b^k}$. Prove that $\sum_{k=0}^{\infty} c_k$ converges and use this to prove that the desired sum converges.

3.7. Tannery's theorem, the exponential function, and the number e

Tannery's theorem (named after Jules Tannery (1848–1910)) is a little known, but fantastic theorem, that I learned from [**31**], [**30**], [**41**], [**73**]. Tannery's theorem is really a special case of the Weierstrass M-test [**41**, p. 124], which is why it probably doesn't get much attention. We shall use Tannery's theorem quite a bit in the sequel. In particular, we shall use it to derive certain properties of the complex exponential function, which is undoubtedly the most important function in analysis and arguably all of mathematics. In this section we derive some of its many properties including its relationship to the number e defined in Section 3.3.

3.7.1. Tannery's theorem for series.

Tannery has two theorems, one for series and the other for products; we'll cover his theorem for products in Section 7.3. Here is the one for series.

Theorem 3.30 (**Tannery's theorem for series**). *For each natural number n, let $\sum_{k=1}^{m_n} a_k(n)$ be a finite sum where $m_n \to \infty$ as $n \to \infty$. If for each k, $\lim_{n\to\infty} a_k(n)$ exists, and there is a series $\sum_{k=1}^{\infty} M_k$ of nonnegative real numbers such that $|a_k(n)| \leq M_k$ for all k, n, then*
$$\lim_{n\to\infty} \sum_{k=1}^{m_n} a_k(n) = \sum_{k=1}^{\infty} \lim_{n\to\infty} a_k(n);$$

that is, both sides are well-defined (the limits and sums converge) and are equal.

Proof. First of all, we remark that the series on the right converges. Indeed, if we put $a_k := \lim_{n\to\infty} a_k(n)$, which exists by assumption, then taking $n \to \infty$ in the inequality $|a_k(n)| \le M_k$, we have $|a_k| \le M_k$ as well. Therefore, by the comparison test, $\sum_{k=1}^\infty a_k$ converges (absolutely).

Now to prove our theorem, let $\varepsilon > 0$ be given. By Cauchy's criterion for series we can fix an ℓ so that

$$M_{\ell+1} + M_{\ell+2} + \cdots < \frac{\varepsilon}{3}.$$

Since $m_n \to \infty$ as $n \to \infty$ we can choose N_1 so that for all $n > N_1$, we have $m_n > \ell$. Then using that $|a_k(n) - a_k| \le |a_k(n)| + |a_k| \le M_k + M_k = 2M_k$, observe that for any $n > N_1$ we have

$$\begin{aligned}\left|\sum_{k=1}^{m_n} a_k(n) - \sum_{k=1}^{\infty} a_k\right| &= \left|\sum_{k=1}^{\ell}(a_k(n) - a_k) + \sum_{k=\ell+1}^{m_n}(a_k(n) - a_k) - \sum_{k=m_n+1}^{\infty} a_k\right| \\ &\le \sum_{k=1}^{\ell}|a_k(n) - a_k| + \sum_{k=\ell+1}^{m_n} 2M_k + \sum_{k=m_n+1}^{\infty} M_k \\ &\le \sum_{k=1}^{\ell}|a_k(n) - a_k| + \sum_{k=\ell+1}^{\infty} 2M_k < \sum_{k=1}^{\ell}|a_k(n) - a_k| + 2\frac{\varepsilon}{3}.\end{aligned}$$

Since for each k, $\lim_{n\to\infty} a_k(n) = a_k$, there is an N such that for each $k = 1, 2, \ldots, \ell$ and for $n > N$, we have $|a_k(n) - a_k| < \varepsilon/(3\ell)$. Thus, if $n > N$, then

$$\left|\sum_{k=1}^{m_n} a_k(n) - \sum_{k=1}^{\infty} a_k\right| < \sum_{k=1}^{\ell}\frac{\varepsilon}{3\ell} + 2\frac{\varepsilon}{3} = \frac{\varepsilon}{3} + 2\frac{\varepsilon}{3}\varepsilon.$$

This completes the proof. □

Tannery's theorem states that under certain conditions we can "switch" limits and *infinite* summations: $\lim_{n\to\infty}\sum_{k=1}^{m_n} a_k(n) = \sum_{k=1}^\infty \lim_{n\to\infty} a_k(n)$. (Of course, we can always switch limits and *finite* summations by the algebra of limits, but infinite summations is a whole other matter.) See Problem 8 for another version of Tannery's theorem and see Problem 9 for an application to double series.

Example 3.45. We shall derive the formula

$$\frac{1}{2} = \lim_{n\to\infty}\left\{\frac{1+2^n}{2^n3+4} + \frac{1+2^n}{2^n3^2+4^2} + \frac{1+2^n}{2^n3^3+4^3} + \cdots + \frac{1+2^n}{2^n3^n+4^n}\right\}.$$

To prove this, we write the right-hand side as

$$\lim_{n\to\infty}\left\{\frac{1+2^n}{2^n3+4} + \frac{1+2^n}{2^n3^2+4^2} + \frac{1+2^n}{2^n3^3+4^3} + \cdots + \frac{1+2^n}{2^n3^n+4^n}\right\} = \lim_{n\to\infty}\sum_{k=1}^{m_n} a_k(n),$$

where $m_n = n$ and

$$a_k(n) := \frac{1+2^n}{2^n3^k+4^k}.$$

Observe that for each $k \in \mathbb{N}$,

$$\lim_{n\to\infty} a_k(n) = \lim_{n\to\infty}\frac{1+2^n}{2^n3^k+4^k} = \lim_{n\to\infty}\frac{\frac{1}{2^n}+1}{3^k+\frac{4^k}{2^n}} = \frac{1}{3^k}$$

exists. Also,

$$|a_k(n)| = \frac{1+2^n}{2^n 3^k + 4^k} \le \frac{2^n+2^n}{2^n 3^k} = \frac{2\cdot 2^n}{2^n 3^k} = \frac{2}{3^k} =: M_k.$$

By the geometric series test, we know that $\sum_{k=1}^\infty M_k$ converges. Hence by Tannery's theorem, we have

$$\begin{aligned}\lim_{n\to\infty} &\left\{ \frac{1+2^n}{2^n 3+4} + \frac{1+2^n}{2^n 3^2+4^2} + \frac{1+2^n}{2^n 3^3+4^3} + \cdots + \frac{1+2^n}{2^n 3^n+4^n} \right\} \\ &= \lim_{n\to\infty} \sum_{k=1}^{m_n} a_k(n) = \sum_{k=1}^{\infty} \lim_{n\to\infty} a_k(n) = \sum_{k=1}^{\infty} \frac{1}{3^k} = \frac{1/3}{1-1/3} = \frac{1}{2}.\end{aligned}$$

If the hypotheses of Tannery's theorem are not met, then the conclusion of Tannery's theorem may not hold as the following example illustrates.

Example 3.46. Here's a non-example of Tannery's theorem. For each $k, n \in \mathbb{N}$, let $a_k(n) := \frac{1}{n}$ and let $m_n = n$. Then

$$\lim_{n\to\infty} a_k(n) = \lim_{n\to\infty} \frac{1}{n} = 0 \quad \Longrightarrow \quad \sum_{k=1}^{\infty} \lim_{n\to\infty} a_k(n) = \sum_{k=1}^{\infty} 0 = 0.$$

On the other hand,

$$\sum_{k=1}^{m_n} a_k(n) = \sum_{k=1}^{n} \frac{1}{n} = \frac{1}{n} \cdot \sum_{k=1}^{n} 1 = 1 \quad \Longrightarrow \quad \lim_{n\to\infty} \sum_{k=1}^{m_n} a_k(n) = \lim_{n\to\infty} 1 = 1.$$

Thus, for this example,

$$\lim_{n\to\infty} \sum_{k=1}^{m_n} a_k(n) \ne \sum_{k=1}^{\infty} \lim_{n\to\infty} a_k(n).$$

What went wrong here is that there is no constant M_k such that $|a_k(n)| \le M_k$ for all n where the series $\sum_{k=1}^\infty M_k$ converges. Indeed, the inequality $|a_k(n)| \le M_k$ for all n implies (setting $n = 1$) that $1 \le M_k$. It follows that the series $\sum_{k=1}^\infty M_k$ cannot converge. Therefore, Tannery's theorem cannot be applied.

3.7.2. The exponential function. The **exponential function** $\exp : \mathbb{C} \to \mathbb{C}$ is the function defined by

$$\boxed{\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad \text{for } \ z \in \mathbb{C}.}$$

Of course, we need to show that the right-hand side converges for each $z \in \mathbb{C}$. In fact, we claim that the series defining $\exp(z)$ is absolutely convergent. To prove this, fix $z \in \mathbb{C}$ and choose $k \in \mathbb{N}$ so that $|z| \le \frac{k}{2}$. (Just as a reminder, recall that such a k exists by the Archimedean property.) Then for any $n \ge k$, we have

$$\begin{aligned}\frac{|z|^n}{n!} &= \left[\left(\frac{|z|}{1}\right)\cdot\left(\frac{|z|}{2}\right)\cdots\left(\frac{|z|}{k}\right)\right]\cdot\left[\left(\frac{|z|}{(k+1)}\right)\cdots\left(\frac{|z|}{n}\right)\right] \\ &\le |z|^k \left(\frac{k}{2(k+1)}\right)\cdot\left(\frac{k}{2(k+2)}\right)\cdots\left(\frac{k}{2n}\right) \\ &\le \left(\frac{k}{2}\right)^k \left(\frac{1}{2}\right)\cdot\left(\frac{1}{2}\right)\cdots\left(\frac{1}{2}\right) = k^k \, \frac{1}{2^n}.\end{aligned}$$

Therefore, for $n \geq k$, $\frac{|z|^n}{n!} \leq \frac{C}{2^n}$ where C is the constant k^k. Since the geometric series $\sum 1/2^n$ converges, by the comparison test, the series defining $\exp(z)$ is absolutely convergent for any $z \in \mathbb{C}$. In the following theorem, we relate the exponential function to Euler's number e introduced in Section 3.3. The proof of Property *(1)* in this theorem is a beautiful application of Tannery's theorem.

Theorem 3.31 (**Properties of the complex exponential**). *The exponential function has the following properties:*

(1) For any $z \in \mathbb{C}$ and sequence $z_n \to z$, we have

$$\exp(z) = \lim_{n\to\infty} \left(1 + \frac{z_n}{n}\right)^n.$$

In particular, setting $z_n = z$ for all n,

$$\exp(z) = \lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n,$$

and setting $z = 1$, we get

$$\exp(1) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e.$$

(2) For any complex numbers z and w,

$$\exp(z) \cdot \exp(w) = \exp(z + w).$$

(3) $\exp(z)$ *is never zero for any complex number z, and*

$$\frac{1}{\exp(z)} = \exp(-z).$$

Proof. To prove *(1)*, let $z \in \mathbb{C}$ and let $\{z_n\}$ be a complex sequence and suppose that $z_n \to z$; we need to show that $\lim_{n\to\infty}(1 + z_n/n)^n = \exp(z)$. To begin, we expand $(1 + z_n/n)^n$ using the binomial theorem:

$$\left(1 + \frac{z_n}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{z_n^k}{n^k} \quad \Longrightarrow \quad \left(1 + \frac{z_n}{n}\right)^n = \sum_{k=0}^{n} a_k(n),$$

where $a_k(n) = \binom{n}{k}\frac{z_n^k}{n^k}$. Hence, we are aiming to prove that

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k(n) = \exp(z).$$

Of course, written in this way, we are in the perfect set-up for Tannery's theorem! However, before going to Tannery's theorem, we note that, by definition of $a_k(n)$, we have $a_0(n) = 1$ and $a_1(n) = z_n$. Therefore, since $z_n \to z$,

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k(n) = \lim_{n\to\infty} \left(1 + z_n + \sum_{k=2}^{n} a_k(n)\right) = 1 + z + \lim_{n\to\infty} \sum_{k=2}^{n} a_k(n).$$

Thus, we just have to apply Tannery's theorem to the sum starting from $k = 2$; for this reason, we henceforth assume that $k, n \geq 2$. Now observe that for $2 \leq k \leq n$, we have

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

Thus, for $2 \le k \le n$,

$$a_k(n) = \frac{1}{k!}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] z_n^k.$$

Using this expression for $a_k(n)$ we can easily verify the hypotheses of Tannery's theorem. First, since $z_n \to z$,

$$\lim_{n\to\infty} a_k(n) = \lim_{n\to\infty} \frac{1}{k!}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] z_n^k = \frac{z^k}{k!}.$$

Second, since $\{z_n\}$ is a convergent sequence, it must be bounded, say by a constant C, so that $|z_n| \le C$ for all n. Then for $2 \le k \le n$,

$$\begin{aligned} |a_k(n)| &= \left|\frac{1}{k!}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] z_n^k\right| \\ &\le \frac{1}{k!}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right] C^k \le \frac{C^k}{k!} =: M_k, \end{aligned}$$

where we used that the term in brackets is product of positive numbers ≤ 1 so the product is also ≤ 1. Note that $\sum_{k=2}^\infty M_k = \sum_{k=2}^\infty C^k/k!$ converges (its sum equals $\exp(C) - 1 - C$, but this isn't important). Hence by Tannery's theorem,

$$\begin{aligned} \lim_{n\to\infty}\left(1+\frac{z_n}{n}\right)^n = 1 + z + \lim_{n\to\infty}\sum_{k=2}^n a_k(n) &= 1 + z + \sum_{k=2}^\infty \lim_{n\to\infty} a_k(n) \\ &= 1 + z + \sum_{k=2}^\infty \frac{z^k}{k!} = \exp(z). \end{aligned}$$

To prove *(2)*, observe that

$$\exp(z)\cdot\exp(w) = \lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n\left(1+\frac{w}{n}\right)^n = \lim_{n\to\infty}\left(1+\frac{z_n}{n}\right)^n,$$

where $z_n = z + w + (z+w)/n$. Since $z_n \to z + w$, we obtain

$$\exp(z)\cdot\exp(w) = \exp(z+w).$$

In particular,

$$\exp(z)\cdot\exp(-z) = \exp(z-z) = \exp(0) = 1,$$

which implies *(3)*. □

We remark that Tannery's theorem can also be used to establish formulas for sine and cosine, see Problem 2. Also, in Section 4.6 we'll see that $\exp(z) = e^z$; however, at this point, we don't even know what e^z ("e to the power z") means.

3.7.3. Approximation and irrationality of e. We now turn to the question of approximating e. Because $n!$ grows very large as $n \to \infty$, we can use the series for the exponential function to calculate e quite easily. If s_n denotes the n-th partial sum for $e = \exp(1)$, then

$$\begin{aligned} e &= s_n + \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \frac{1}{(n+3)!} + \cdots \\ &= s_n + \frac{1}{(n+1)!} + \frac{1}{(n+1)!(n+2)} + \frac{1}{(n+1)!(n+2)(n+3)} + \cdots \\ &< s_n + \frac{1}{(n+1)!}\left\{1 + \frac{1}{(n+1)} + \frac{1}{(n+1)^2} + \cdots\right\} \\ &= s_n + \frac{1}{(n+1)!} \cdot \frac{1}{1 - \cfrac{1}{n+1}} = s_n + \frac{1}{n!\,n}. \end{aligned}$$

Thus, we get the following useful estimate for e:

$$\boxed{s_n < e < s_n + \frac{1}{n!\,n}.} \tag{3.42}$$

Example 3.47. In particular, with $n = 1$ we have $s_1 = 2$ and $1/(1!\,1) = 1$, therefore $2 < e < 3$. Of course, we can get a much more precise estimate with higher values of n: with $n = 10$ we obtain (in common decimal notation — see Section 3.8)

$$2.718281801 < e < 2.718281829.$$

Thus, only $n = 10$ gives a quite accurate approximation!

The estimate (3.42) also gives an easy proof that the number e is irrational, a fact first proved by Euler in 1737 [**36**, p. 463].

Theorem 3.32 (**Irrationality of** e). *e is irrational.*

Proof. Indeed, by way of contradiction suppose that $e = p/q$ where p and q are positive integers with $q > 1$. Then (3.42) with $n = q$ implies that

$$s_q < \frac{p}{q} < s_q + \frac{1}{q!\,q}.$$

Since $s_q = 2 + \frac{1}{2!} + \cdots + \frac{1}{q!}$, the number $q!\,s_q$ is an integer (this is because $q! = 1 \cdot 2 \cdots k \cdot (k+1) \cdots q$ contains a factor of $k!$ for each $1 \leq k \leq q$), which we denote by m. Then multiplying the inequalities $s_q < \frac{p}{q} < s_q + \frac{1}{q!\,q}$ by $q!$ and using the fact that $q > 1$, we obtain

$$m < p\,(q-1)! < m + \frac{1}{q} < m + 1.$$

Hence the integer $p\,(q-1)!$ lies between m and $m+1$, which of course is absurd, since there is no integer between m and $m+1$. □

We end with the following neat "infinite nested product" formula for e:

$$\boxed{e = 1 + \frac{1}{1} + \frac{1}{2}\Big(1 + \frac{1}{3}\Big(1 + \frac{1}{4}\Big(1 + \frac{1}{5}\Big(\ \cdots\ \Big)\Big)\Big)\Big);} \tag{3.43}$$

see Problem 6 for the meaning of the right-hand side.

Exercises 3.7.

1. Determine the following limits.

$$(a)\ \lim_{n\to\infty}\left\{\frac{1+n}{(1+2n)}+\frac{2^2+n^2}{(1+2n)^2}+\cdots+\frac{n^n+n^n}{(1+2n)^n}\right\},$$

$$(b)\ \lim_{n\to\infty}\left\{\frac{n}{\sqrt{1+(1\cdot 2\cdot n)^2}}+\frac{n}{\sqrt{1+(2\cdot 3\cdot n)^2}}+\cdots+\frac{n}{\sqrt{1+(n\cdot(n+1)\cdot n)^2}}\right\},$$

$$(c)\ \lim_{n\to\infty}\left\{\frac{1+n^2}{1+(1\cdot n)^2}+\frac{2^2+n^2}{1+(2\cdot n)^2}+\cdots+\frac{n^2+n^2}{1+(n\cdot n)^2}\right\},$$

$$(d)\ \lim_{n\to\infty}\left\{\left(\frac{n}{1+1^{2n}}\right)^{\frac{1}{n}}+\left(\frac{n}{1+2^{2n}}\right)^{\frac{1}{n}}+\cdots+\left(\frac{n}{1+n^{2n}}\right)^{\frac{1}{n}}\right\},$$

where for (c) and (d), prove that the limits are $\sum_{k=1}^\infty \frac{1}{k^2}$.

2. For each $z\in\mathbb{C}$, define the **cosine** of z by

$$\cos z := \lim_{n\to\infty}\frac{1}{2}\left\{\left(1+\frac{iz}{n}\right)^n+\left(1-\frac{iz}{n}\right)^n\right\}$$

and the **sine** of z by

$$\sin z := \lim_{n\to\infty}\frac{1}{2i}\left\{\left(1+\frac{iz}{n}\right)^n-\left(1-\frac{iz}{n}\right)^n\right\}.$$

(a) Use Tannery's theorem in a similar way as we did in the proof of Property *(1)* in Theorem 3.31 to prove that the limits defining $\cos z$ and $\sin z$ exists and moreover,

$$\cos z=\sum_{k=0}^\infty(-1)^k\frac{z^{2k}}{(2k)!}\quad\text{and}\quad \sin z=\sum_{k=0}^\infty(-1)^k\frac{z^{2k+1}}{(2k+1)!}.$$

(b) Following the proof that e is irrational, prove that $\cos 1$ (or $\sin 1$ if you prefer) is irrational.

3. Following [**139**], we prove that for any $m\geq 3$,

$$\sum_{n=0}^m\frac{1}{n!}-\frac{3}{2\,m}<\left(1+\frac{1}{m}\right)^m<\sum_{n=0}^m\frac{1}{n!}.$$

Taking $m\to\infty$ gives an alternative proof that $\exp(1)=e$. Fix $m\geq 3$.

(i) Prove that for any $2\leq k\leq m$, we have

$$1-\frac{k(k-1)}{2m}=1-\frac{(1+2+\cdots+k-1)}{m}\leq\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{k-1}{m}\right)<1.$$

(ii) Using (i), prove that

$$\sum_{n=0}^m\frac{1}{n!}-\frac{1}{2\,m}\sum_{n=0}^{m-2}\frac{1}{n!}<\left(1+\frac{1}{m}\right)^m<\sum_{n=0}^m\frac{1}{n!}.$$

Now prove the formula. Suggestion: Use the binomial theorem on $(1+\frac{1}{m})^m$.

4. Let $\{a_n\}$ be any sequence of rational numbers tending to $+\infty$, that is, given any $M>0$ there is an N such that for all $n>N$, we have $a_n>M$. In this problem we show that

$$e=\lim\left(1+\frac{1}{a_n}\right)^{a_n}. \tag{3.44}$$

This formula also holds when the a_n's are real numbers, but as of now, we only know about rational powers (we'll consider real powers in Section 4.6).

(i) Prove that (3.44) holds in case the a_n's are integers tending to $+\infty$.

(ii) By the tails theorem, we may assume that $1 < a_n$ for all n. For each n, let m_n be the unique integer such that $m_n - 1 \leq a_n < m_n$ (thus, $m_n = \lfloor a_n \rfloor - 1$ where $\lfloor a_n \rfloor$ is the greatest integer function). Prove that if $m_n \geq 1$, then

$$\left(1 + \frac{1}{m_n}\right)^{m_n - 1} \leq \left(1 + \frac{1}{a_n}\right)^{a_n} \leq \left(1 + \frac{1}{m_n - 1}\right)^{m_n}.$$

Now prove (3.44).

5. Let $\{b_n\}$ be any null sequence of positive rational numbers. Prove that

$$e = \lim \left(1 + b_n\right)^{\frac{1}{b_n}}.$$

6. Prove that for any $n \in \mathbb{N}$,

$$1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} = 1 + \frac{1}{1} + \frac{1}{2}\Big(1 + \frac{1}{3}\Big(1 + \frac{1}{4}\Big(\ \cdots\ \Big(1 + \frac{1}{n-1}\Big(1 + \frac{1}{n}\Big)\Big)\Big)\Big)\Big).$$

The infinite nested sum in (3.43) denotes the limit as $n \to \infty$ of this expression.

7. Trying to imitate the proof that e is irrational, prove that for any $m \in \mathbb{N}$, $\exp(1/m)$ is irrational. After doing this, show that $\cos(1/m)$ (or $\sin(1/m)$ if you prefer) is irrational, where cosine and sine are defined in Problem 2. See the article [**177**] for more on irrationality proofs.

8. (**Tannery's theorem II**) For each natural number n, let $\sum_{k=1}^\infty a_k(n)$ be a convergent series. Prove that if for each k, $\lim_{n\to\infty} a_k(n)$ exists, and there is a series $\sum_{k=1}^\infty M_k$ of nonnegative real numbers such that $|a_k(n)| \leq M_k$ for all k, n, then

$$\lim_{n\to\infty} \sum_{k=1}^\infty a_k(n) = \sum_{k=1}^\infty \lim_{n\to\infty} a_k(n).$$

Suggestion: Try to imitate the proof of the original Tannery's theorem.

9. (**Tannery's theorem and Cauchy's double series theorem**— see Section 6.5 for more on double series!)) In this problem we relate Tannery's theorem to double series. A **double sequence** is just a map $a : \mathbb{N} \times \mathbb{N} \to \mathbb{C}$; for $m, n \in \mathbb{N}$ we denote $a(m,n)$ by a_{mn}. We say that the **iterated series** $\sum_{m=1}^\infty \sum_{n=1}^\infty a_{mn}$ converges if for each $m \in \mathbb{N}$, the series $\sum_{n=1}^\infty a_{mn}$ converges (call the sum α_m) and the series $\sum_{m=1}^\infty \alpha_m$ converges. Similarly, we say that the iterated series $\sum_{n=1}^\infty \sum_{m=1}^\infty a_{mn}$ converges if for each $n \in \mathbb{N}$, the series $\sum_{m=1}^\infty a_{mn}$ converges (call the sum β_n) and the series $\sum_{n=1}^\infty \beta_n$ converges.

The object of this problem is to prove that given any double sequence $\{a_{mn}\}$ of complex numbers such that either

$$\sum_{m=1}^\infty \sum_{n=1}^\infty |a_{mn}| \text{ converges} \qquad \text{or} \qquad \sum_{n=1}^\infty \sum_{m=1}^\infty |a_{mn}| \text{ converges} \tag{3.45}$$

then

$$\sum_{m=1}^\infty \sum_{n=1}^\infty a_{mn} = \sum_{n=1}^\infty \sum_{m=1}^\infty a_{mn} \tag{3.46}$$

in the sense that both iterated sums converge and are equal. The implication (3.45) $\Longrightarrow$ (3.46) is called **Cauchy's double series theorem**; see Theorem 6.26 in Section 6.5 for the full story. To prove this, you may proceed as follows.

(i) Assume that $\sum_{m=1}^\infty \sum_{n=1}^\infty |a_{mn}|$ converges; we must prove the equality (3.46). To do so, for each $k \in \mathbb{N}$ define $M_k := \sum_{j=1}^\infty |a_{kj}|$, which converges by assumption. Then $\sum_{k=1}^\infty M_k$ also converges by assumption. Define $a_k(n) := \sum_{j=1}^n a_{kj}$. Prove that Tannery's theorem II can be applied to these $a_k(n)$'s and in doing so, establish the equality (3.46).

(ii) Assume that $\sum_{n=1}^\infty \sum_{m=1}^\infty |a_{mn}|$ converges; prove the equality (3.46).

(iii) Cauchy's double series theorem can be used to prove neat and non-obvious identities. For example, prove that for any $k \in \mathbb{N}$ and $z \in \mathbb{C}$ with $|z| < 1$, we have

$$\sum_{n=1}^{\infty} \frac{z^{n(k+1)}}{1-z^n} = \sum_{m=1}^{\infty} \frac{z^{m+k}}{1-z^{m+k}};$$

that is,

$$\frac{z^{k+1}}{1-z} + \frac{z^{2(k+1)}}{1-z^2} + \frac{z^{3(k+1)}}{1-z^3} + \cdots = \frac{z^{1+k}}{1-z^{1+k}} + \frac{z^{2+k}}{1-z^{2+k}} + \frac{z^{3+k}}{1-z^{3+k}} + \cdots .$$

Suggestion: Apply Cauchy's double series to $\{a_{mn}\}$ where $a_{mn} = z^{n(m+k)}$.

3.8. Decimals and "most" numbers are transcendental á la Cantor

Since grade school we have represented real numbers in "base 10".[7] In this section we continue our discussion initiated in Section 2.5 (for integers) on the use of arbitrary bases for real numbers. We also look at "Cantor's diagonal argument" that is able to *construct* transcendental numbers.

3.8.1. Decimal and b-adic representations of real numbers. We are all familiar with the common decimal or base 10 notation, which we used without mention in the last section concerning the estimate $2.718281801 < e < 2.718281829$. Here, we know that the decimal (also called base 10) *notation* 2.718281801 represents the *number*

$$2 + \frac{7}{10} + \frac{1}{10^2} + \frac{8}{10^3} + \frac{2}{10^4} + \frac{8}{10^5} + \frac{1}{10^6} + \frac{8}{10^7} + \frac{0}{10^8} + \frac{1}{10^9},$$

that is, this real number gives meaning to the symbol 2.718281801. More generally, the *symbol* $\alpha_k\alpha_{k-1}\cdots\alpha_0.a_1a_2a_3a_4a_5\ldots$, where the α_n's and a_n's are integers in $0, 1, \ldots, 9$, represents the *number*

$$\alpha_k \cdot 10^k + \cdots + \alpha_1 \cdot 10 + \alpha_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots = \sum_{n=0}^{k} \alpha_n \cdot 10^n + \sum_{n=1}^{\infty} \frac{a_n}{10^n}.$$

Notice that the infinite series $\sum_{n=1}^{\infty} \frac{a_n}{10^n}$ converges because $0 \le a_n \le 9$ for all n so we can compare this series with $\sum_{n=1}^{\infty} \frac{9}{10^n} = 1 < \infty$. In particular, the number $\sum_{n=1}^{\infty} \frac{a_n}{10^n}$ lies in $[0, 1]$.

More generally, instead of restricting to base 10, we can use other bases. Let $b > 1$ be an integer (the **base**). Then the *symbol* $\alpha_k\alpha_{k-1}\cdots\alpha_0 ._b a_1a_2a_3a_4a_5\ldots$, where the α_n's and a_n's are integers in $0, 1, \ldots, b-1$, represents the real *number*

$$a = \alpha_k \cdot b^k + \cdots + \alpha_1 \cdot b + \alpha_0 + \frac{a_1}{b} + \frac{a_2}{b^2} + \frac{a_3}{b^3} + \cdots = \sum_{n=0}^{k} \alpha_n \cdot b^n + \sum_{n=1}^{\infty} \frac{a_n}{b^n}.$$

The infinite series $\sum_{n=1}^{\infty} \frac{a_n}{b^n}$ converges because $0 \le a_n \le b-1$ for all n so we can compare this series with

$$\sum_{n=1}^{\infty} \frac{b-1}{b^n} = (b-1) \cdot \sum_{n=1}^{\infty} \left(\frac{1}{b}\right)^n = (b-1) \frac{\frac{1}{b}}{1-\frac{1}{b}} = 1 < \infty.$$

[7] *To what heights would science now be raised if Archimedes had made that discovery ! [= the decimal system of numeration or its equivalent (with some base other than 10)]. Carl Friedrich Gauss (1777–1855).*

The symbol $\alpha_k\alpha_{k-1}\cdots\alpha_0 \cdot_b a_1a_2a_3a_4a_5\ldots$ is called the **b-adic representation** or **b-adic expansion** of a. The numbers α_n and a_n are called **digits**. A natural question is: Does every real number have such a representation? The answer is yes.

Recall from Section 2.5 that integers have b-adic representations so we can focus on noninteger numbers. Now given any $x \in \mathbb{R}\setminus\mathbb{Z}$, we can write $x = m + a$ where $m = \lfloor x \rfloor$ is the integer part of x and $0 < a < 1$. We already know that m has a b-adic expansion, so we can focus on writing a in a b-adic expansion.

THEOREM 3.33. *Let $b \in \mathbb{N}$ with $b > 1$. Then for any $a \in [0,1]$, there exists a sequence of integers $\{a_n\}_{n=1}^\infty$ with $0 \le a_n \le b-1$ for all n such that*

$$a = \sum_{n=1}^\infty \frac{a_n}{b^n};$$

if $a \neq 0$, then infinitely many of the a_n's are nonzero.

PROOF. If $a = 0$ we must (and can) take all the a_n's to be zero, so we may assume that $a \in (0,1]$; we find the a_n's as follows. First, we divide $(0,1]$ into b disjoint intervals:

$$\Big(0, \frac{1}{b}\Big],\ \Big(\frac{1}{b}, \frac{2}{b}\Big],\ \Big(\frac{2}{b}, \frac{3}{b}\Big], \ldots, \Big(\frac{b-1}{b}, 1\Big].$$

Since $a \in (0,1]$, a must lie in one of these intervals, so there is an integer a_1 with $0 \le a_1 \le b-1$ such that

$$a \in \Big(\frac{a_1}{b}, \frac{a_1+1}{b}\Big] \quad\Longleftrightarrow\quad \frac{a_1}{b} < a \le \frac{a_1+1}{b}.$$

Second, we divide $\big(\frac{a_1}{b}, \frac{a_1+1}{b}\big]$ into b disjoint subintervals. Since the length of $\big(\frac{a_1}{b}, \frac{a_1+1}{b}\big]$ is $\frac{a_1+1}{b} - \frac{a_1}{b} = \frac{1}{b}$, we divide the interval $\big(\frac{a_1}{b}, \frac{a_1+1}{b}\big]$ into b subintervals of length $(1/b)/b = 1/b^2$:

$$\Big(\frac{a_1}{b}, \frac{a_1}{b} + \frac{1}{b^2}\Big], \Big(\frac{a_1}{b} + \frac{1}{b^2}, \frac{a_1}{b} + \frac{2}{b^2}\Big], \Big(\frac{a_1}{b} + \frac{2}{b^2}, \frac{a_1}{b} + \frac{3}{b^2}\Big], \ldots, \Big(\frac{a_1}{b} + \frac{b-1}{b^2}, \frac{a_1}{b} + \frac{1}{b}\Big].$$

Now $a \in \big(\frac{a_1}{b}, \frac{a_1+1}{b}\big]$, so a must lie in one of these intervals. Thus, there is an integer a_2 with $0 \le a_2 \le b-1$ such that

$$a \in \Big(\frac{a_1}{b} + \frac{a_2}{b^2}, \frac{a_1}{b} + \frac{a_2+1}{b^2}\Big] \quad\Longleftrightarrow\quad \frac{a_1}{b} + \frac{a_2}{b^2} < a \le \frac{a_1}{b} + \frac{a_2+1}{b^2}.$$

Continuing this process (slang for "by induction") we can find a sequence of integers $\{a_n\}$ such that $0 \le a_n \le b-1$ for all n and

$$\frac{a_1}{b} + \frac{a_2}{b^2} + \cdots + \frac{a_{n-1}}{b^{n-1}} + \frac{a_n}{b^n} < a \le \frac{a_1}{b} + \cdots + \frac{a_{n-1}}{b^{n-1}} + \frac{a_n+1}{b^n}. \tag{3.47}$$

Let $y := \sum_{n=1}^\infty \frac{a_n}{b^n}$; this series converges because its partial sums are bounded by a according to the left-hand inequality in (3.47). Since $1/b^n \to 0$ as $n \to \infty$ by taking $n \to \infty$ in (3.47) and using the squeeze rule, we see that $y \le a \le y$. This shows that $a = y = \sum_{n=1}^\infty \frac{a_n}{b^n}$. There must be infinitely many nonzero a_n's for if there were only finitely many nonzero a_n's, say for some m we have $a_n = 0$ for all $n > m$, then we would have $a = \sum_{n=1}^m \frac{a_n}{b^n}$. Now setting $n = m$ in (3.47) and looking at the left-hand inequality shows that $a < a$. This is impossible, so there must be infinitely many nonzero a_n's. □

Here's another question: If a b-adic representation exists, is it unique? The answer to this question is no.

Example 3.48. Consider, for example, the number 1/2, which has two decimal expansions:

$$\frac{1}{2} = 0.50000000\ldots \quad \text{and} \quad \frac{1}{2} = 0.49999999\ldots.$$

Notice that the first decimal expansion terminates.

You might remember from high school that the only decimals with two different expansions are the ones that terminate. In general, a b-adic expansion $0._b a_1a_2a_3a_4a_5\ldots$ is said to **terminate** if all the a_n's equal zero for n large.

Theorem 3.34. *Let b be a positive integer greater than 1. Then every real number in $(0,1]$ has a unique b-adic expansion, except a terminating expansion, which also can have a b-adic expansion where $a_n = b-1$ for all n sufficiently large.*

Proof. For $a \in (0,1]$, let $a = \sum_{n=1}^\infty \frac{a_n}{b^n}$ be its b-adic expansion found in Theorem 3.33, so there are infinitely many nonzero a_n's. Suppose that $\{\alpha_n\}$ is another sequence of integers, not equal to the sequence $\{a_n\}$, such that $0 \le \alpha_n \le b-1$ for all n and such that $a = \sum_{n=1}^\infty \frac{\alpha_n}{b^n}$. Since $\{a_n\}$ and $\{\alpha_n\}$ are not the same sequence there is at least one n such that $a_n \neq \alpha_n$. Let m be the smallest natural number such that $a_m \neq \alpha_m$. Then $a_n = \alpha_n$ for $n = 1, 2, \ldots, m-1$, so

$$\sum_{n=1}^\infty \frac{a_n}{b^n} = \sum_{n=1}^\infty \frac{\alpha_n}{b^n} \quad \Longrightarrow \quad \sum_{n=m}^\infty \frac{a_n}{b^n} = \sum_{n=m}^\infty \frac{\alpha_n}{b^n}.$$

Since there are infinitely many nonzero a_n's, we have

$$\begin{aligned} \frac{a_m}{b^m} < \sum_{n=m}^\infty \frac{a_n}{b^n} = \sum_{n=m}^\infty \frac{\alpha_n}{b^n} &= \frac{\alpha_m}{b^m} + \sum_{n=m+1}^\infty \frac{\alpha_n}{b^n} \\ &\le \frac{\alpha_m}{b^m} + \sum_{n=m+1}^\infty \frac{b-1}{b^n} = \frac{\alpha_m}{b^m} + \frac{1}{b^m}. \end{aligned}$$

Multiplying the extremities of these inequalities by b^m, we obtain $a_m < \alpha_m + 1$, so $a_m \le \alpha_m$. Since we know that $a_m \neq \alpha_m$, we must actually have $a_m < \alpha_m$. Now

$$\begin{aligned} \frac{\alpha_m}{b^m} \le \sum_{n=m}^\infty \frac{\alpha_n}{b^n} = \sum_{n=m}^\infty \frac{a_n}{b^n} &= \frac{a_m}{b^m} + \sum_{n=m+1}^\infty \frac{a_n}{b^n} \\ &\le \frac{a_m}{b^m} + \sum_{n=m+1}^\infty \frac{b-1}{b^n} = \frac{a_m}{b^m} + \frac{1}{b^m} \le \frac{\alpha_m}{b^m}. \end{aligned}$$

Since the ends are equal, all the inequalities in between must be equalities. In particular, making the first inequality into an equality shows that $\alpha_n = 0$ for all $n = m+1, m+2, m+3, \ldots$ and making the middle inequality into an equality shows that $a_n = b-1$ for all $n = m+1, m+2, m+3, \ldots$. It follows that a has only one b-adic expansion except when we can write a as

$$a = 0._b \alpha_1\alpha_2\ldots\alpha_m = 0._b a_1\ldots a_m(b-1)(b-1)(b-1)\ldots,$$

a terminating one, and one that has repeating $b-1$'s. This completes our proof. □

3.8.2. Rational numbers. We now consider periodic decimals, such as

$$\frac{1}{3} = 0.3333333\ldots, \quad \frac{3526}{495} = 7.1232323\ldots, \quad \frac{611}{495} = 1.2343434\ldots.$$

As you well know, we usually write these decimals as

$$\frac{1}{3} = 0.\overline{3}\ldots, \quad \frac{3526}{495} = 7.1\overline{23}\ldots, \quad \frac{611}{495} = 1.2\overline{34}\ldots.$$

For general b-adic expansions, we say that $\alpha_k \ldots \alpha_0 ._b a_1 a_2 a_3 \cdots$ is **periodic** if there exists an $\ell \in \mathbb{N}$ (called a **period**) such that $a_n = a_{n+\ell}$ for all n sufficiently large.

Example 3.49. For example, in the base 10 expansion of $\frac{3526}{495} = 7.1232323\ldots = 7.a_1 a_2 a_3 \ldots$, we have $a_n = a_{n+2}$ for all $n \geq \ell = 2$.

We can actually see how the periodic pattern appears by going back to high school long division! Indeed, long dividing 495 into 3526 we get

$$\begin{array}{r}
7.123\\
495\,\overline{)3526.000}\\
\underline{3465}\\
610\\
\underline{495}\\
1150\\
\underline{990}\\
1600\\
\underline{1485}\\
115
\end{array}$$

At this point, we get another remainder of 115, exactly as we did a few lines before. Thus, by continuing this process of long division, we are going to repeat the pattern $2, 3$. We shall use this long division technique to prove the following theorem.

THEOREM 3.35. *Let b be a positive integer greater than* 1*. A real number is rational if and only if its b-adic expansion is periodic.*

PROOF. We first prove the "only if", then the "if" statement.

Step 1: We prove the "only if": Given integers p, q with $q > 0$, we show that p/q has a periodic b-adic expansion. By the division algorithm (see Theorem 2.15), we can write $p/q = q' + r/q$ where $q' \in \mathbb{Z}$ and $0 \leq r < q$. Thus, we just have to prove that r/q has a periodic b-adic expansion. In particular, we might as well assume from the beginning that $0 < p < q$ so that $p/q < 1$. Proceeding via high school long division, we construct the decimal expansion of p/q.

First, using the division algorithm, we divide bp by q, obtaining a unique integer a_1 such that $bp = a_1 q + r_1$ where $0 \leq r_1 < q$. Since

$$\frac{p}{q} - \frac{a_1}{b} = \frac{bp - a_1 q}{bq} = \frac{r_1}{bq} \geq 0,$$

we have

$$\frac{a_1}{b} \leq \frac{p}{q} < 1,$$

which implies that $0 \leq a_1 < b$.

Next, using the division algorithm, we divide $b r_1$ by q, obtaining a unique integer a_2 such that $br_1 = a_2 q + r_2$ where $0 \leq r_2 < q$. Since

$$\frac{p}{q} - \frac{a_1}{b} - \frac{a_2}{b^2} = \frac{r_1}{bq} - \frac{a_2}{b^2} = \frac{br_1 - a_2 q}{b^2 q} = \frac{r_2}{b^2 q} \geq 0,$$

we have

$$\frac{a_2}{b^2} \leq \frac{r_1}{bq} < \frac{q}{bq} = \frac{1}{b},$$

which implies that $0 \leq a_2 < b$.

Once more using the division algorithm, we divide $b\,r_2$ by q, obtaining a unique integer a_3 such that $br_2 = a_3q + r_3$ where $0 \leq r_3 < q$. Since

$$\frac{p}{q} - \frac{a_1}{b} - \frac{a_2}{b^2} - \frac{a_3}{b^3} = \frac{r_2}{b^2q} - \frac{a_3}{b^3} = \frac{br_2 - a_3q}{b^3q} = \frac{r_3}{b^3q} \geq 0,$$

we have

$$\frac{a_3}{b^3} \leq \frac{r_2}{b^2q} < \frac{q}{b^2q} = \frac{1}{b^2},$$

which implies that $0 \leq a_3 < b$. Continuing by induction, we construct integers $0 \leq a_n, r_n < q$ such that for each n, $br_n = a_{n+1}q + r_{n+1}$ and

$$\frac{p}{q} - \frac{a_1}{b} - \frac{a_2}{b^2} - \frac{a_3}{b^3} - \cdots - \frac{a_n}{b^n} = \frac{r_n}{b^nq}.$$

Since $0 \leq r_n < q$ it follows that $\frac{r_n}{b^nq} \to 0$ as $n \to \infty$, so we can write

$$\frac{p}{q} = \sum_{n=1}^{\infty} \frac{a_n}{b^n} \quad \Longleftrightarrow \quad \frac{p}{q} = 0._b a_1a_2a_3a_4a_5 \ldots. \tag{3.48}$$

Now one of two things holds: Either some remainder $r_n = 0$ or none of the r_n's are zero. Suppose that we are in the first case, some $r_n = 0$. By construction, we divide br_n by q using the division algorithm to get $br_n = a_{n+1}q + r_{n+1}$. Since $r_n = 0$ and quotients and remainders are unique, we must have $a_{n+1} = 0$ and $r_{n+1} = 0$. By construction, we divide br_{n+1} by q using the division algorithm to get $br_{n+1} = a_{n+2}q + r_{n+2}$. Since $r_{n+1} = 0$ and quotients and remainders are unique, we must have $a_{n+2} = 0$ and $r_{n+2} = 0$. Continuing this procedure, we see that all a_k with $k > n$ are zero. This, in view of (3.48), shows that the b-adic expansion of p/q has repeating zeros, so in particular is periodic.

Suppose that we are in the second case, no $r_n = 0$. Consider the $q + 1$ remainders $r_1, r_2, \ldots, r_{q+1}$. Since $0 \leq r_n < q$, each r_n can only take on the q values $0, 1, 2, \ldots, q - 1$ ("q holes"), so by the pigeonhole principle, two of these remainders must have the same value ("be in the same hole"). Thus, $r_k = r_{k+\ell}$ for some k and ℓ. We now show that $a_{k+1} = a_{k+\ell+1}$. Indeed, a_{k+1} was defined by dividing br_k by q so that $br_k = a_{k+1}q + r_{k+1}$. On the other hand, $a_{k+\ell+1}$ was defined by dividing $br_{k+\ell}$ by q so that $br_{k+\ell} = a_{k+\ell+1}q + r_{k+\ell+1}$. Now the division algorithm states that the quotients and remainders are unique. Since $br_k = br_{k+\ell}$, it follows that $a_{k+1} = a_{k+\ell+1}$ and $r_{k+1} = r_{k+\ell+1}$. Repeating this same argument shows that $a_{k+n} = a_{k+\ell+n}$ for all $n \geq 0$; that is, $a_n = a_{n+\ell}$ for all $n \geq k$. Thus, p/q has a periodic b-adic expansion.

Step 2: We now prove the "if" portion: A number with a periodic b-adic expansion is rational. Let a be a real number and suppose that its b-adic decimal expansion is periodic. Since a is rational if and only of its noninteger part is rational, we may assume that the integer part of a is zero. Let

$$a = 0._b a_1a_2 \cdots a_k\overline{b_1 \cdots b_\ell}$$

have a periodic b-adic expansion, where the bar means that the block $b_1 \cdots b_\ell$ repeats. Observe that in an expansion $\alpha_m\alpha_{m-1} \cdots \alpha_0._b\beta_1\beta_2\beta_3 \ldots$, multiplication by

b^n for $n \in \mathbb{N}$ moves the decimal point n places to the right. (Try to prove this; think about the familiar base 10 case first.) In particular,

$$b^{k+\ell}a = a_1a_2\cdots a_kb_1\cdots b_\ell ._b\overline{b_1\cdots b_\ell} = a_1a_2\cdots a_kb_1\cdots b_\ell + 0._b\overline{b_1\cdots b_\ell}$$

and

$$b^k a = a_1a_2\cdots a_k ._b\overline{b_1\cdots b_\ell} = a_1a_2\cdots a_k + 0._b\overline{b_1\cdots b_\ell}.$$

Subtracting, we see that the numbers given by $0._b\overline{b_1\cdots b_\ell}$ cancel, so $b^{k+\ell}a - b^k a = p$, where p is an integer. Hence, $a = p/q$, where $q = b^{k+\ell} - b^k$. Thus a is rational. □

3.8.3. Cantor's diagonal argument. Now that we know about decimal expansions, we can present Cantor's second proof that the real numbers are uncountable. His first proof appeared in Section 2.10.

THEOREM 3.36 (**Cantor's second proof**). *The interval* $(0,1)$ *is uncountable.*

PROOF. Assume, for sake of deriving a contradiction, that there is a bijection $f : \mathbb{N} \longrightarrow (0,1)$. Let us write the images of f as decimals (base 10):

$$\begin{aligned}
1 &\longleftrightarrow f(1) = .a_{11}\, a_{12}\, a_{13}\, a_{14} \cdots \\
2 &\longleftrightarrow f(2) = .a_{21}\, a_{22}\, a_{23}\, a_{24} \cdots \\
3 &\longleftrightarrow f(3) = .a_{31}\, a_{32}\, a_{33}\, a_{34} \cdots \\
4 &\longleftrightarrow f(4) = .a_{41}\, a_{42}\, a_{43}\, a_{44} \cdots \\
&\vdots \qquad\quad \vdots,
\end{aligned}$$

where we may assume that in each of these expansions there is never an infinite run of 9's. Recall from Theorem 3.33 there every real number of $(0,1)$ has a *unique* such representation. Now let us define a real number $a = .a_1\, a_2\, a_3\, \cdots$, where

$$a_n := \begin{cases} 3 & \text{if } a_{nn} \neq 3 \\ 7 & \text{if } a_{nn} = 3. \end{cases}$$

(The choice of 3 are 7 is arbitrary — you can choose another pair of unequal integers in $0,\ldots,9$ if you like!) Notice that $a_n \neq a_{nn}$ for all n. In particular, $a \neq f(1)$ because a and $f(1)$ differ in the first digit. On the other hand, $a \neq f(2)$ because a and $f(2)$ differ in the second digit. Similarly, $a \neq f(n)$ for every n since a and $f(n)$ differ in the n-th digit. This contradicts that $f : \mathbb{N} \to (0,1)$ is onto. □

This argument is not only elegant, it is useful: Cantor's diagonal argument gives a good method to *generate* transcendental numbers (see [**87**])!

EXERCISES 3.8.

1. Find the numbers with the b-adic expansions (here $b = 10, 2, 3$, respectively):
$$(a)\ 0.010101\ldots, \quad (b)\ 0._2 010101\ldots, \quad (c)\ 0._3 010101\ldots.$$
2. Prove that a real number $a \in (0,1)$ has a terminating decimal expansion if and only if $2^m 5^n a \in \mathbb{Z}$ for some nonnegative integers m, n.
3. (s**-adic expansions**) Let $s = \{b_n\}$ be a sequence of integers with $b_n > 1$ for all n and let $0 < a \le 1$. Prove that there is a sequence of integers $\{a_n\}_{n=1}^\infty$ with $0 \le a_n \le b_n - 1$ for all n and with infinitely many nonzero a_n's such that
$$a = \sum_{n=1}^\infty \frac{a_n}{b_1 \cdot b_2 \cdot b_3 \cdots b_n},$$
Suggestion: Can you imitate the proof of Theorem 3.33?

4. (**Cantor's original diagonal argument**) Let g and c be any two distinct objects and let G be the set consisting of all functions $f : \mathbb{N} \longrightarrow \{g, c\}$. Let $f_1, f_2, f_3, \ldots$ be any infinite sequence of elements of G. Prove that there is an element f in G that is not in this list. From this prove that G is uncountable. Conclude that the set of all sequences of 0's and 1's is uncountable.

CHAPTER 4

Limits, continuity, and elementary functions

> *One merit of mathematics few will deny: it says more in fewer words than any other science. The formula, $e^{i\pi} = -1$ expressed a world of thought, of truth, of poetry, and of the religious spirit "God eternally geometrizes."*
> *David Eugene Smith (1860–1944)* [**188**].

In this chapter we study, without doubt, the most important types of functions in all of analysis and topology, the continuous functions. In particular, we study the continuity properties of the "the most important function in mathematics" [**192**, p. 1]: $\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$, $z \in \mathbb{C}$. From this single function arise just about every important function and number you can think of: the logarithm function, powers, roots, the trigonometric functions, the hyperbolic functions, the number e, the number π, ..., and the famous formula displayed in the above quote!

What do the Holy Bible, squaring the circle, House bill No. 246 of the Indiana state legislature in 1897, square free natural numbers, coprime natural numbers, the sentence

(4.1) *May I have a large container of coffee? Thank you,*

the mathematicians Archimedes of Syracuse, William Jones, Leonhard Euler, Johann Heinrich Lambert, Carl Louis Ferdinand von Lindemann, John Machin, and Yasumasa Kanada have to do with each other? The answer (drum role please): They all have been involved in the life of the remarkable number π! This fascinating number is defined and some of its amazing and death-defying properties and formulæ are studied in this chapter! By the way, the sentence (4.1) is a mnemonic device to remember the digits of π. The number of letters in each word represents a digit of π; e.g. "May" represents 3, "I" 1, etc. The sentence (4.1) gives ten digits of π: 3.141592653.[1]

In Section 4.1 we begin our study of continuity by learning limits of functions, in Section 4.2 we study some useful limit properties, and then in Section 4.3 we discuss continuous functions in terms of limits of functions. In Section 4.4, we study some fundamental properties of continuous functions. A special class of functions, called monotone functions, have many special properties, which are investigated in Section 4.5. In Section 4.6 we study "the most important function in mathematics" and we also study its inverse, the logarithm function, and then we use the logarithm function to define powers. We also define the Riemann zeta function, the Euler-Mascheroni constant γ:

$$\boxed{\gamma := \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right),}$$

[1] Using mnemonics to memorize digits of π isn't a good idea if you want to beat Hiroyuki Goto's record of reciting 42,195 digits from memory! (see http://www.pi-world-ranking-list.com)

a constant will come up again and again (see the book [**96**], which is devoted to this number), and we'll prove that the alternating harmonic series has sum log 2:

$$\boxed{\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + - \cdots\,,}$$

another fact that will come up often. In Section 4.7 we use the exponential function to define the trigonometric functions and we define π, the fundamental constant of geometry. In Section 4.8 we study roots of complex numbers and we give fairly elementary proofs of the fundamental theorem of algebra. In Section 4.9 we study the inverse trigonometric functions. The calculation and (hopeful) imparting of a sense of great fascination of the incredible number π are the features of Sections 4.10, 5.1 and 5.2. In particular, we'll derive the first analytical expression for π:

$$\boxed{\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots\,,}$$

given in 1593 by François Viète (1540–1603) [**47**, p. 69], Gregory-Leibniz-Madhava's formula for $\pi/4$:

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + - \cdots\,,}$$

and Euler's solution to the famous Basel problem. Here, the Basel problem was the following: Find the sum of the reciprocals of the squares of the natural numbers, $\sum_{n=1}^{\infty} \frac{1}{n^2}$; the answer, first given by Euler in 1734, is $\pi^2/6$:

$$\boxed{\frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \cdots\,.}$$

Chapter 4 objectives: The student will be able to ...

- apply the rigorous ε-δ definition of limits for functions and continuity.
- apply and understand the proofs of the fundamental theorem of continuous functions.
- define the elementary functions (exponential, trigonometric, and their inverses) and the number π.
- explain three related proofs of the fundamental theorem of algebra.

4.1. Convergence and ε-δ arguments for limits of functions

In elementary calculus you most likely studied limits of functions without rigorous proofs, using intuition, graphs, or informal reasoning to determine limits. In this section we define limits and seek to truly understand them *precisely.*

4.1.1. Limit points and the ε-δ definition of limit. Before reading on, it might benefit the reader to reread the material on open balls in Section 2.8. If $A \subseteq \mathbb{R}^m$, then a point $c \in \mathbb{R}^m$ is said to be a **limit point** of A if every open ball centered at c contains a point of A different from c. In other words, given any $r > 0$, there is a point $x \in A$ such that $x \in B_r(c)$ and $x \neq c$, which is to say,

$$\boxed{c \text{ is a limit point of } A \iff \text{for each } r > 0, \text{ there's an } x \in A \text{ with } 0 < |x - c| < r.}$$

The inequality $0 < |x - c|$ just means that $x \neq c$ while the inequality $|x - c| < r$ just means that $x \in B_r(c)$. If $m = 1$, then c is a limit point of A if for any $r > 0$, there is a point $x \in A$ such that $x \in (c - r, c + r)$ and $x \neq c$. We remark that the point c may or may not belong to A.

Example 4.1. Let $A = [0, 1)$. Then 0 is a limit point of A and 1 is also a limit point of A; in this example, 0 belongs to A while 1 does not. Moreover, as the reader can verify, the set of all limit points of A is the closed interval $[0, 1]$.

Example 4.2. If $A = \{1/n\,;\, n \in \mathbb{N}\}$, then the diligent reader will verify that 0 is the only limit point of A. (Note that every open ball centered at 0 contains a point in A by the $1/n$-principle.)

The name "limit point" fits because the following lemma states that limit points are exactly that, limits of points in A.

LEMMA 4.1 (**Limit points and sequences**). *A point $c \in \mathbb{R}^m$ is a limit point of a set $A \subseteq \mathbb{R}^m$ if and only if $c = \lim a_n$ for some sequence $\{a_n\}$ contained in A with $a_n \neq c$ for each n.*

PROOF. Assume that c is a limit point of A. For each n, by definition of limit point (put $r = 1/n$), there is a point $a_n \in A$ such that $0 < |a_n - c| < 1/n$. We leave the reader to check that $a_n \to c$.

Conversely, suppose that $c = \lim a_n$ for some sequence $\{a_n\}$ contained in A with $a_n \neq c$ for each n. We shall prove that c is a limit point for A. Let $r > 0$. Then by definition of convergence for $a_n \to c$, there is an n sufficiently large such that $|a_n - c| < r$. Since $a_n \neq c$ by assumption, we have $0 < |x - c| < r$ with $x = a_n \in A$, so c is a limit point of A. □

We now define limits of functions. Let $m, p \in \mathbb{N}$ and let $f : D \longrightarrow \mathbb{R}^m$ where $D \subseteq \mathbb{R}^p$. From elementary calculus, we learn that $L = \lim_{x \to c} f(x)$ indicates that $f(x)$ is "as close as we want" to L for $x \in D$ "sufficiently close", but not equal, to c. We now make the terms in quotes rigorous. As with limits of sequences, we interpret "as close as we want" to mean that given any error $\varepsilon > 0$, for $x \in D$ "sufficiently close", but not equal, to c we can approximate L by $f(x)$ to within an error of ε. In other words, for $x \in D$ "sufficiently close", but not equal, to c, we have

$$|f(x) - L| < \varepsilon.$$

We interpret "sufficiently close" to mean that there is a real number $\delta > 0$ such that for all $x \in D$ with $|x - c| < \delta$ and $x \neq c$, the above inequality holds; since $x \neq c$ we have $|x - c| > 0$, so we are in effect saying $0 < |x - c| < \delta$. In summary, for all $x \in D$ with $0 < |x - c| < \delta$ we have $|f(x) - L| < \varepsilon$.

We now conclude our findings as a precise definition. A function $f : D \longrightarrow \mathbb{R}^m$ is said to have a **limit** L at a limit point c of D if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\boxed{x \in D \quad \text{and} \quad 0 < |x - c| < \delta \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.} \tag{4.2}$$

(Note that since c is a limit point of D there always exists points $x \in D$ with $0 < |x - c| < \delta$, so this implication is not an empty implication.) If this holds, we write

$$L = \lim_{x \to c} f \quad \text{or} \quad L = \lim_{x \to c} f(x),$$

or sometimes expressed by $f \to L$ or $f(x) \to L$ as $x \to c$. Of course, we can use another letter instead of "x" to denote the domain variable.

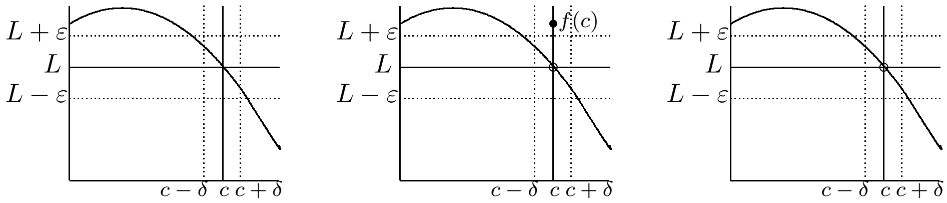


FIGURE 4.1. Here are three functions with $D = [0,\infty)$ (we'll denote them by the generic letter f). In the first graph, $L = f(c)$, in the second graph $f(c) \neq L$, and in the third graph, $f(c)$ is not even defined. However, in all three cases, $\lim_{x\to c} f = L$.

An alternative definition of limit involves open balls. A function $f : D \longrightarrow \mathbb{R}^m$ has limit L at a limit point c of D if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x \in D \cap B_\delta(c) \quad \text{and} \quad x \neq c \quad \Longrightarrow \quad f(x) \in B_\varepsilon(L).$$

For $p = m = 1$, this condition simplifies as follows: $f : D \longrightarrow \mathbb{R}$ has limit L at a limit point c of $D \subseteq \mathbb{R}$ if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x \in D \cap (c-\delta, c+\delta) \quad \text{and} \quad x \neq c \quad \Longrightarrow \quad f(x) \in (L-\varepsilon, L+\varepsilon),$$

which is to say (see Figure 4.1 for an illustration of this limit concept),

$$x \in D \ \text{ with } \ c-\delta < x < c+\delta \ \text{ and } \ x \neq c \quad \Longrightarrow \quad L-\varepsilon < f(x) < L+\varepsilon.$$

We take the convention that if not explicitly mentioned, the domain D of a function is always taken to be the set of all points for which the function makes sense.

4.1.2. Working with the ε-δ definition. Here are some examples to master.

Example 4.3. Let us prove that

$$\lim_{x\to 2} \big(3x^2 - 10\big) = 2.$$

(Here, the domain D of $3x^2 - 10$ is assumed to be all of $\mathbb{R}$.) Let $\varepsilon > 0$ be given. We need to prove that there is a real number $\delta > 0$ such that

$$0 < |x-2| < \delta \quad \Longrightarrow \quad \big|3x^2 - 10 - 2\big| = \big|3x^2 - 12\big| < \varepsilon.$$

How do we find such a δ ... well ... we "massage" $|3x^2 - 12|$. Observe that

$$\big|3x^2 - 12\big| = 3\,\big|x^2 - 4\big| = 3\,|x+2| \cdot |x-2|.$$

Let us tentatively restrict x so that $|x-2| < 1$. In this case,

$$|x+2| = |x-2+4| \leq |x-2| + 4 < 1 + 4 = 5.$$

Thus,

$$(4.3) \qquad |x-2| < 1 \quad \Longrightarrow \quad \big|3x^2 - 12\big| = 3\,|x+2| \cdot |x-2| < 15\,|x-2|.$$

Now

$$(4.4) \qquad 15\,|x-2| < \varepsilon \quad \Longleftrightarrow \quad |x-2| < \frac{\varepsilon}{15}.$$

For this reason, let us pick δ to be the minimum of 1 and $\varepsilon/15$. Then $|x-2| < \delta$ implies $|x-2| < 1$ and $|x-2| < \varepsilon/15$, therefore according to (4.3) and (4.4), we have

$$0 < |x-2| < \delta \quad \Longrightarrow \quad \big|3x^2 - 12\big| \overset{\text{by (4.3)}}{<} 15\,|x-2| \overset{\text{by (4.4)}}{<} \varepsilon.$$

Thus, by definition of limit, $\lim_{x\to 2}(3x^2 - 10) = 2$.

Example 4.4. Now let $a > 0$ be any real number and let us show that

$$\lim_{x\to 0} \frac{\sqrt{x+a} - \sqrt{a}}{x} = \frac{1}{2\sqrt{a}}.$$

(Here, the domain $D = [-a, 0) \cup (0, \infty)$.) Let $\varepsilon > 0$ be any given positive real number. We need to prove that there is a real number $\delta > 0$ such that

$$0 < |x| < \delta \quad \Longrightarrow \quad \left| \frac{\sqrt{x+a} - \sqrt{a}}{x} - \frac{1}{2\sqrt{a}} \right| < \varepsilon.$$

To establish this result we "massage" the absolute value with the "multiply by conjugate trick":

$$(4.5) \qquad \sqrt{x+a} - \sqrt{a} = \frac{\sqrt{x+a} - \sqrt{a}}{1} \cdot \frac{\sqrt{x+a} + \sqrt{a}}{\sqrt{x+a} + \sqrt{a}} = \frac{x}{\sqrt{x+a} + \sqrt{a}}.$$

Therefore,

$$\left| \frac{\sqrt{x+a} - \sqrt{a}}{x} - \frac{1}{2\sqrt{a}} \right| = \left| \frac{1}{\sqrt{x+a} + \sqrt{a}} - \frac{1}{2\sqrt{a}} \right| = \left| \frac{\sqrt{x+a} - \sqrt{a}}{2\sqrt{a}\,(\sqrt{x+a} + \sqrt{a})} \right|.$$

Applying (4.5) to the far right numerator, we get

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

Observe that $(\sqrt{x+a} + \sqrt{a})^2 \geq (0 + \sqrt{a})^2 = a$, so

$$\frac{1}{2\sqrt{a}\,(\sqrt{x+a} + \sqrt{a})^2} \leq \frac{1}{2\sqrt{a}\cdot a} = \frac{1}{2a^{3/2}} \quad \Longrightarrow \quad \left| \frac{\sqrt{x+a} - \sqrt{a}}{x} - \frac{1}{2\sqrt{a}} \right| \leq \frac{|x|}{2a^{3/2}},$$

such a simple expression! Now

$$\frac{|x|}{2a^{3/2}} < \varepsilon \quad \Longleftrightarrow \quad |x| < 2a^{3/2}\,\varepsilon.$$

With this in mind, we choose $\delta = 2a^{3/2}\,\varepsilon$ and with this choice of δ, we obtain our desired inequality:

$$x \in D \quad \text{and} \quad 0 < |x| < \delta \quad \Longrightarrow \quad \left| \frac{\sqrt{x+a} - \sqrt{a}}{x} - \frac{1}{2\sqrt{a}} \right| < \varepsilon.$$

Example 4.5. Here is an example involving complex numbers. Let c be any nonzero complex number and let us show that

$$\lim_{z\to c} \frac{1}{z} = \frac{1}{c}.$$

Here, $f : D \longrightarrow \mathbb{C}$ is the function $f(z) = 1/z$ with $D \subseteq \mathbb{C}$ consisting of all nonzero complex numbers. (Recall that $\mathbb{C} = \mathbb{R}^2$, so D is a subset of $\mathbb{R}^2$ and in terms of our original definition (4.2), $D \subseteq \mathbb{R}^p$ and $f : D \longrightarrow \mathbb{R}^m$ with $p = m = 2$.) Let $\varepsilon > 0$ be any given positive real number. We need to prove that there is a real number $\delta > 0$ such that

$$0 < |z - c| < \delta \quad \Longrightarrow \quad \left| \frac{1}{z} - \frac{1}{c} \right| < \varepsilon.$$

Now

$$\left| \frac{1}{z} - \frac{1}{c} \right| = \left| \frac{c - z}{zc} \right| = \frac{1}{|zc|}\,|z - c|.$$

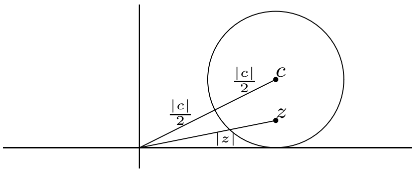


FIGURE 4.2. If $|z-c|<\frac{|c|}{2}$, then this picture shows that $|z|>\frac{|c|}{2}$.

In order to make this expression less than ε, we need to bound the term in front of $|z-c|$ (we need to make sure that $|z|$ can't get too small, otherwise $1/|zc|$ can blow-up). To do so, we tentatively restrict z so that $|z-c|<\frac{|c|}{2}$. In this case, as seen in Figure 4.2, we also have $|z|>\frac{|c|}{2}$. Here is a proof if you like:

$$|c|=|c-z+z|\le|c-z|+|z|<\frac{|c|}{2}+|z| \quad\Longrightarrow\quad \frac{|c|}{2}<|z|.$$

Therefore, if $|z-c|<\frac{|c|}{2}$, then $|zc|>\frac{|c|}{2}\cdot|c|=\frac{1}{2}|c|^2=b$, where $b=|c|^2/2$ is a positive number. Thus,

$$|z-c|<\frac{|c|}{2} \quad\Longrightarrow\quad \left|\frac{1}{z}-\frac{1}{c}\right|<\frac{1}{b}\,|z-c|. \tag{4.6}$$

Now

$$\frac{1}{b}\,|z-c|<\varepsilon \quad\Longleftrightarrow\quad |z-c|<b\,\varepsilon. \tag{4.7}$$

For this reason, let us pick δ to be the minimum of $|c|/2$ and $b\,\varepsilon$. Then $|z-c|<\delta$ implies $|z-c|<|c|/2$ and $|z-c|<b\,\varepsilon$, therefore according to (4.6) and (4.7), we have

$$0<|z-c|<\delta \quad\Longrightarrow\quad \left|\frac{1}{z}-\frac{1}{c}\right| \overset{\text{by (4.6)}}{<} \frac{1}{b}\,|z-c| \overset{\text{by (4.7)}}{<} \varepsilon.$$

Thus, by definition of limit, $\lim_{z\to c}1/z=1/c$.

Example 4.6. Here is one last example. Define $f:\mathbb{R}^2\setminus\{0\}\longrightarrow\mathbb{R}$ by

$$f(x)=\frac{x_1^2\,x_2}{x_1^2+x_2^2}\ , \quad x=(x_1,x_2).$$

We shall prove that $\lim_{x\to 0}f=0$. (In the subscript "$x\to 0$", 0 denotes the zero vector $(0,0)$ in $\mathbb{R}^2$ while on the right of $\lim_{x\to 0}f=0$, 0 denotes the real number 0; it should always be clear from context what "0" means.) Before our actual proof, we first note that for any real numbers a,b, we have $0\le(a-b)^2=a^2+b^2-2ab$. Solving for ab, we get

$$\boxed{a\,b\le\frac{1}{2}\big(a^2+b^2\big).}$$

This inequality is well worth remembering. Hence,

$$|f(x_1,x_2)|=\frac{|x_1|}{x_1^2+x_2^2}\cdot|x_1\,x_2|\le\frac{|x_1|}{x_1^2+x_2^2}\cdot\frac{1}{2}\big(x_1^2+x_2^2\big)=\frac{|x_1|}{2}.$$

Given $\varepsilon>0$, choose $\delta=\varepsilon$. Then

$$0<|x|<\delta \quad\Longrightarrow\quad |x_1|<\delta \quad\Longrightarrow\quad |f(x)|\le\frac{|x_1|}{2}\le\frac{\varepsilon}{2}<\varepsilon,$$

which implies that $\lim_{x\to 0}f=0$.

4.1.3. The sequence definition of limit. It turns out that we can relate limits of functions to limits of sequences, which was studied in Chapter 3, so we can use much of the theory developed in that chapter to analyze limits of functions. In particular, take note of the following important theorem!

Theorem 4.2 (**Sequence criterion for limits**). *Let $f : D \longrightarrow \mathbb{R}^m$ and let c be a limit point of D. Then $L = \lim_{x\to c} f$ if and only if for every sequence $\{a_n\}$ of points in $D \setminus \{c\}$ with $c = \lim a_n$, we have $L = \lim_{n\to\infty} f(a_n)$.*

Proof. Let $f : D \longrightarrow \mathbb{R}^m$ and let c be a limit point of D. We first prove that if $L = \lim_{x\to c} f$, then for any sequence $\{a_n\}$ of points in $D \setminus \{c\}$ converging to c, we have $L = \lim f(a_n)$. To do so, let $\{a_n\}$ be such a sequence and let $\varepsilon > 0$. Since f has limit L at c, there is a $\delta > 0$ such that

$$x \in D \quad \text{and} \quad 0 < |x - c| < \delta \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.$$

Since $a_n \to c$ and $a_n \neq c$ for any n, it follows that there is an N such that

$$n > N \quad \Longrightarrow \quad 0 < |a_n - c| < \delta.$$

The limit property of f now implies that

$$n > N \quad \Longrightarrow \quad |f(a_n) - L| < \varepsilon.$$

Thus, $L = \lim f(a_n)$.

We now prove that if for every sequence $\{a_n\}$ of points in $D \setminus \{c\}$ converging to c, we have $L = \lim f(a_n)$, then $L = \lim_{x\to c} f$. We prove the logically equivalent contrapositive; that is, if $L \neq \lim_{x\to c} f$, then there is a sequence $\{a_n\}$ of points in $D\setminus\{c\}$ converging to c such that $L \neq \lim f(a_n)$. Now $L \neq \lim_{x\to c} f$ means (negating the definition $L = \lim_{x\to c} f$) that there is an $\varepsilon > 0$ such that for all $\delta > 0$, there is an $x \in D$ with $0 < |x - c| < \delta$ and $|f(x) - L| \geq \varepsilon$. Since this statement is true for all $\delta > 0$, it is in particular true for $\delta = 1/n$ for each $n \in \mathbb{N}$. Thus, for each $n \in \mathbb{N}$, there is a point $a_n \in D$ with $0 < |a_n - c| < 1/n$ and $|f(a_n) - L| \geq \varepsilon$. It follows that $\{a_n\}$ is a sequence of points in $D \setminus \{c\}$ converging to c and $\{f(a_n)\}$ does not converge to L. This completes the proof of the contrapositive. □

This theorem can be used to prove that certain functions don't have limits.

Example 4.7. Recall from Section 1.3 the **Dirichlet function**, named after Johann Peter Gustav Lejeune Dirichlet (1805–1859):

$$D : \mathbb{R} \longrightarrow \mathbb{R} \quad \text{is defined by} \quad D(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Let $c \in \mathbb{R}$. Then as we saw in Example 3.13 of Section 3.2, there is a sequence $\{a_n\}$ of rational numbers converging to c with $a_n \neq c$ for all n. Since a_n is rational we have $D(a_n) = 1$ for all $n \in \mathbb{N}$, so

$$\lim D(a_n) = \lim 1 = 1.$$

Also, there is a sequence $\{b_n\}$ of irrational numbers converging to c with $b_n \neq c$ for all n, in which case

$$\lim D(b_n) = \lim 0 = 0.$$

Therefore, according to our sequence criterion, $\lim_{x\to c} D$ cannot exist.

Example 4.8. Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$f(x) = \frac{x_1 x_2}{x_1^2 + x_2^2}, \qquad x = (x_1, x_2) \neq 0.$$

We claim that $\lim_{x\to 0} f$ does not exist. To see this, observe that $f(x_1, 0) = 0$, so $f(a_n) \to 0$ for any sequence a_n approaching 0 along the x_1-axis. On the other hand, since

$$f(x_1, x_1) = \frac{x_1^2}{x_1^2 + x_1^2} = \frac{1}{2},$$

it follows that $f(a_n) \to 1/2$ for any sequence a_n approaching 0 along the diagonal $x_1 = x_2$. Therefore $\lim_{x\to 0} f$ does not exist.

Exercises 4.1.

1. Using the ε-δ definition of limit, prove that (where z is a complex variable)

$$(a)\ \lim_{z\to 1} \left(z^2 + 2z\right) = 3, \quad (b)\ \lim_{z\to 2} z^3 = 8, \quad (c)\ \lim_{z\to 2} \frac{1}{z^2} = \frac{1}{4}, \quad (d)\ \lim_{z\to 2} \frac{3z}{z+1} = 2,.$$

Suggestion: For (b), can you factor $z^3 - 8$?

2. Using the ε-δ definition of limit, prove that (where x, a are real variables and where in (b) and (c), $a > 0$)

$$(a)\ \lim_{x\to a} \frac{x^2 - x - a^2}{x + a} = -\frac{1}{2}, \quad (b)\ \lim_{x\to a} \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{a}}, \quad (c)\ \lim_{x\to 0} \frac{\sqrt{a^2 + 6x^2} - a}{x^2} = \frac{3}{a}.$$

3. Prove that the limits (a) and (b) do not exist while (c) does exist:

$$(a)\ \lim_{x\to 0} \frac{x_1^2 + x_2}{\sqrt{x_1^2 + x_2^2}}, \quad (b)\ \lim_{x\to 0} \frac{x_1^2 + x_2}{x_1^2 + x_2^2}, \quad (c)\ \lim_{x\to 0} \frac{x_1^2 x_2^2}{x_1^2 + x_2^2}.$$

4. Here are problems involving functions similar to Dirichlet's function. Define

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}, \qquad g(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 1 - x & \text{if } x \text{ is irrational.} \end{cases}$$

(a) Prove that $\lim_{x\to 0} f = 0$, but $\lim_{x\to c} f$ does not exist for $c \neq 0$.

(b) Prove that $\lim_{x\to 1/2} g = 1/2$, but $\lim_{x\to c} g$ does not exist for $c \neq 1/2$.

5. Let $f : D \longrightarrow \mathbb{R}$ with $D \subseteq \mathbb{R}^p$ and let $L = \lim_{x\to c} f$. Assume that $f(x) \geq 0$ for all $x \neq c$ sufficiently close to c. Prove that $L \geq 0$ and $\sqrt{L} = \lim_{x\to c} \sqrt{f(x)}$.

4.2. A potpourri of limit properties for functions

Now that we have a working knowledge of the ε-δ definition of limit for functions, we move onto studying the properties of limits that will be used throughout the rest of the book.

4.2.1. Limit theorems.

As we already mentioned in Section 4.1.3, combining the sequence criterion for limits (Theorem 4.2) with the limit theorems in Chapter 3, we can easily prove results concerning limits. Here are some examples beginning with the following companion to the uniqueness theorem (Theorem 3.2) for sequences.

Theorem 4.3 (**Uniqueness of limits**). *A function can have at most one limit at any given limit point of its domain.*

Proof. If $\lim_{x\to c} f$ equals both L and L', then according to the sequence criterion, for all sequences $\{a_n\}$ of points in $D \setminus \{c\}$ converging to c, we have $\lim f(a_n) = L$ and $\lim f(a_n) = L'$. Since we know that limits of sequences are unique (Theorem 3.2), we conclude that $L = L'$. □

If $f : D \longrightarrow \mathbb{R}^m$, then we can write f in terms of its components as

$$f = (f_1, \ldots, f_m),$$

where for $k = 1, 2, \ldots, m$, $f_k : D \longrightarrow \mathbb{R}$, are the **component functions** of f. In particular, if $f : D \longrightarrow \mathbb{C}$, then we can always break up f as

$$f = (f_1, f_2) \quad \Longleftrightarrow \quad f = f_1 + i f_2,$$

where $f_1, f_2 : D \longrightarrow \mathbb{R}$.

Example 4.9. For instance, if $f : \mathbb{C} \longrightarrow \mathbb{C}$ is defined by $f(z) = z^2$, then we can write this as $f(x + iy) = (x + iy)^2 = x^2 - y^2 + i2xy$. Therefore, $f = f_1 + i f_2$ where if $z = x + iy$, then

$$f_1(z) = x^2 - y^2, \;\; f_2(z) = 2xy.$$

The following theorem is a companion to the component theorem (Theorem 3.4) for sequences.

Theorem 4.4 (**Component theorem**). *A function converges to $L \in \mathbb{R}^m$ (at a given limit point of the domain) if and only if each component function converges in $\mathbb{R}$ to the corresponding component of L.*

Proof. Let $f : D \longrightarrow \mathbb{R}^m$. Then $\lim_{x\to c} f = L$ if and only if for every sequence $\{a_n\}$ of points in $D \setminus \{c\}$ converging to c, we have $\lim f(a_n) = L$. According to the component theorem for sequences, $\lim f(a_n) = L$ if and only if for each $k = 1, 2, \ldots, m$, we have $\lim_{n\to\infty} f_k(a_n) = L_k$. This shows that $\lim_{x\to c} f = L$ if and only if for each $k = 1, 2, \ldots, m$, $\lim_{x\to c} f_k = L_k$ and completes our proof. □

The following theorem is a function analog of the "algebra of limits" studied in Section 3.2.

Theorem 4.5 (**Algebra of limits**). *If f and g both have limits as $x \to c$, then*

(1) $\lim_{x\to c} |f| = |\lim_{x\to c} f|$.

(2) $\lim_{x\to c} \big(af + bg\big) = a \lim_{x\to c} f + b \lim_{x\to c} g$, *for any real a, b.*

If f and g take values in $\mathbb{C}$, then

(3) $\lim_{x\to c} fg = \big(\lim_{x\to c} f\big)\big(\lim_{x\to c} g\big)$.

(4) $\lim_{x\to c} f/g = \lim_{x\to c} f / \lim_{x\to c} g$, *provided the denominators are not zero for x near c.*

Proof. All these properties follow from the corresponding statements for sequences in Theorems 3.10 and 3.11. For example, let us prove *(4)* and leave the rest to the reader. If $L = \lim_{x\to c} f$ and $L' = \lim_{x\to c} g$, then by the sequence criterion (Theorem 4.2) it suffices to show that for any sequence $\{a_n\}$ of points in $D\setminus\{c\}$ converging to c, we have $\lim f(a_n)/g(a_n) = L/L'$. However, given any such sequence, by the sequence criterion we know that $\lim f(a_n) = L$ and $\lim g(a_n) = L'$, and by Theorem 3.11, we thus have $\lim f(a_n)/g(a_n) = (\lim f(a_n))/(\lim g(a_n)) = L/L'$. □

By induction, we can use the algebra of limits on any finite sum or finite product of functions.

Example 4.10. It is easy to show that at any point $c \in \mathbb{C}$, $\lim_{z\to c} z = c$. Therefore, by our algebra of limits, for any complex number a and natural number n, we have

$$\lim_{z\to c} az^n = a \lim_{z\to c} \underbrace{z \cdot z \cdots z}_{n\ z\text{'s}} = a\left(\lim_{z\to c} z\right) \cdot \left(\lim_{z\to c} z\right) \cdots \left(\lim_{z\to c} z\right) = a\, c \cdot c \cdots c = a\, c^n.$$

Therefore, given any polynomial

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

for any $c \in \mathbb{C}$, the algebra of limits implies that

$$\begin{aligned} \lim_{z\to c} p(z) &= \lim_{z\to c} a_n\, z^n + \lim_{z\to c} a_{n-1}\, z^{n-1} + \cdots + \lim_{z\to c} a_1 z + \lim_{z\to c} a_0 \\ &= a_n\, c^n + a_{n-1}\, c^{n-1} + \cdots + a_1\, c + a_0; \end{aligned}$$

that is,

$$\lim_{z\to c} p(z) = p(c).$$

Example 4.11. Now let $q(z)$ be another polynomial and suppose that $q(c) \neq 0$. Since $q(z)$ has at most finitely many roots (Proposition 2.53), it follows that $q(z) \neq 0$ for z sufficiently close to c. Therefore by our algebra of limits, we have

$$\lim_{z\to c} \frac{p(z)}{q(z)} = \frac{\lim_{z\to c} p(z)}{\lim_{z\to c} q(z)} = \frac{p(c)}{q(c)}.$$

The following theorem is useful when dealing with compositions of functions.

Theorem 4.6 (**Composition of limits**). *Let $f : D \longrightarrow \mathbb{R}^m$ and $g : C \longrightarrow \mathbb{R}^p$ where $D \subseteq \mathbb{R}^p$ and $C \subseteq \mathbb{R}^q$ and suppose that $g(C) \subseteq D$ so that $f \circ g : C \longrightarrow \mathbb{R}^m$ is defined. Let d be a limit point of D and c a limit point of C and assume that*

(1) $d = \lim_{x\to c} g(x)$.
(2) $L = \lim_{y\to d} f(y)$.
(3) Either $f(d) = L$ or $d \neq g(x)$ for all $x \neq c$ sufficiently near c.

Then

$$L = \lim_{x\to c} f \circ g.$$

Proof. Let $\{a_n\}$ be any sequence in $C \setminus \{c\}$ converging to c. Then by *(1)*, the sequence $\{g(a_n)\}$ in D converges to d.

We now consider the two cases in *(3)*. First, If $g(x) \neq d$ for all $x \neq c$ sufficiently near c, then a tail of the sequence $\{g(a_n)\}$ is a sequence in $D \setminus \{d\}$ converging to d, so by *(2)*, $\lim f(g(a_n)) = L$. On the other hand, if it is the case that $f(d) = L$ then by *(2)* and the definition of limit it follows that for *any* sequence $\{b_n\}$ in D converging to d, we have $\lim f(b_n) = L$. Therefore, in this case we also have $\lim f(g(a_n)) = L$. In either case, we get $L = \lim_{x\to c} f \circ g$. □

We now finish our limit theorems by considering limits and inequalities. In the following two theorems, all functions map a subset $D \subseteq \mathbb{R}^p$ into $\mathbb{R}$.

Theorem 4.7 (**Squeeze theorem**). *Let f, g, and h be such that $f(x) \leq g(x) \leq h(x)$ for all x sufficiently close to a limit point c in D and such that both limits $\lim_{x\to c} f$ and $\lim_{x\to c} h$ exist and are equal. Then the limit $\lim_{x\to c} g$ also exists, and*

$$\lim_{x\to c} f = \lim_{x\to c} g = \lim_{x\to c} h.$$

As with the previous two theorems, the squeeze theorem for functions is a direct consequence of the sequence criterion and the corresponding squeeze theorem for sequences (Theorem 3.7) and therefore we shall omit the proof. The next theorem follows (as you might have guessed) the sequence criterion and the corresponding preservation of inequalities theorem (Theorem 3.8) for sequences.

THEOREM 4.8 (**Preservation of inequalities**). *Suppose that* $\lim_{x\to c} f$ *exists.*

(1) If $\lim_{x\to c} g$ *exists and* $f(x) \leq g(x)$ *for* $x \neq c$ *sufficiently close to* c*, then* $\lim_{x\to c} f \leq \lim_{x\to c} g$*.*

(2) If for some real numbers a *and* b*, we have* $a \leq f(x) \leq b$ *for* $x \neq c$ *sufficiently close to* c*, then* $a \leq \lim_{x\to c} f \leq b$*.*

4.2.2. Limits, limits, limits, and more limits. When the domain is a subset of $\mathbb{R}$, there are various extensions of the limit idea. We begin with left and right-hand limits. For the rest of this section we consider functions $f : D \longrightarrow \mathbb{R}^m$ where $D \subseteq \mathbb{R}$ (later we'll further restrict to $m = 1$).

Suppose that c is a limit point of the set $D \cap (-\infty, c)$. Then $f : D \longrightarrow \mathbb{R}^m$ is said to have a **left-hand limit** L at c if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\boxed{x \in D \quad \text{and} \quad c - \delta < x < c \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.} \tag{4.8}$$

In a similar way we define a right-hand limit: Suppose that c is a limit point of the set $D \cap (c, \infty)$. Then f is said to have a **right-hand limit** L at c if for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\boxed{x \in D \quad \text{and} \quad c < x < c + \delta \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.} \tag{4.9}$$

We express left-hand limits in one of several ways:

$$L = \lim_{x\to c-} f \ , \quad L = \lim_{x\to c-} f(x) \ , \quad L = f(c-) \ , \quad \text{or } f(x) \to L \text{ as } x \to c-;$$

with similar expressions with $c+$ replacing $c-$ for right-hand limits. An important fact relating one-sided limits and regular limits is described in the next result, whose proof we leave to you.

THEOREM 4.9. *Let* $f : D \longrightarrow \mathbb{R}^m$ *with* $D \subseteq \mathbb{R}$ *and suppose that* c *is a limit point of the sets* $D \cap (-\infty, c)$ *and* $D \cap (c, \infty)$*. Then*

$$L = \lim_{x\to c} f \qquad \Longleftrightarrow \qquad L = f(c-) \ \textit{ and } \ L = f(c+).$$

If only one of $f(c-)$ or $f(c+)$ makes sense, then $L = \lim_{x\to c} f$ if and only if $L = f(c-)$ (when c is only a limit point of $D \cap (-\infty, c)$) or $L = f(c+)$ (when c is only a limit point of $D \cap (c, \infty)$), whichever makes sense.

We now describe limits at infinity. Suppose that for any real number N there is a point $x \in D$ such that $x > N$. A function $f : D \longrightarrow \mathbb{R}^m$ is said to have a **limit** $L \in \mathbb{R}^m$ as $x \to \infty$ if for each $\varepsilon > 0$ there is a $N \in \mathbb{R}$ such that

$$\boxed{x \in D \quad \text{and} \quad x > N \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.} \tag{4.10}$$

Now suppose that for any real number N there is a point $x \in D$ such that $x < N$. A function $f : D \longrightarrow \mathbb{R}^m$ is said to have a **limit** $L \in \mathbb{R}^m$ as $x \to -\infty$ if for each $\varepsilon > 0$ there is a $N \in \mathbb{R}$ such that

$$\boxed{x \in D \quad \text{and} \quad x < N \quad \Longrightarrow \quad |f(x) - L| < \varepsilon.} \tag{4.11}$$

To express these limits at infinity, we use the notations (sometimes with ∞ replaced by $+\infty$)

$$L = \lim_{x\to\infty} f \ , \quad L = \lim_{x\to\infty} f(x) \ , \quad f \to L \text{ as } x \to \infty \ , \quad \text{or } f(x) \to L \text{ as } x \to \infty;$$

with similar expressions when $x \to -\infty$.

Finally, we discuss infinite limits, which are also called properly divergent limits of functions.[2] We now let $m = 1$ and consider functions $f : D \longrightarrow \mathbb{R}$ with $D \subseteq \mathbb{R}$. Suppose that for any real number N there is a point $x \in D$ such that $x > N$. Then f is said to **diverge to** ∞ as $x \to \infty$ if for any real number $M > 0$ there is a $N \in \mathbb{R}$ such that

$$\boxed{x \in D \quad \text{and} \quad x > N \quad \Longrightarrow \quad M < f(x).} \tag{4.12}$$

Also, f is said to **diverge to** $-\infty$ as $x \to \infty$ if for any real number $M < 0$ there is a $N \in \mathbb{R}$ such that

$$\boxed{x \in D \quad \text{and} \quad x > N \quad \Longrightarrow \quad f(x) < M.} \tag{4.13}$$

In either case we say that f is **properly divergent** as $x \to \infty$ and when f is properly divergent to ∞ we write

$$\infty = \lim_{x\to\infty} f \ , \quad \infty = \lim_{x\to\infty} f(x) \ , \quad f \to \infty \text{ as } x \to \infty \ , \quad \text{or } f(x) \to \infty \text{ as } x \to \infty;$$

with similar expressions when f properly diverges to $-\infty$. *In a very similar manner we can define properly divergent limits of functions as $x \to -\infty$, as $x \to c$, as $x \to c-$, and $x \to c+$; we leave these other definitions for the reader to formulate.*

Let us now consider an example.

Example 4.12. Let $a > 1$ and let $f : \mathbb{Q} \longrightarrow \mathbb{R}$ be defined by $f(x) = a^x$ (therefore in this case, $D = \mathbb{Q}$). Here, we recall that a^x is defined for any rational number x (see Section 2.7). We shall prove that

$$\lim_{x\to\infty} f = \infty \quad \text{and} \quad \lim_{x\to-\infty} f = 0.$$

In Section 4.6 we shall define a^x for any $x \in \mathbb{R}$ (in fact, for any complex power) and we shall establish these same limits with $D = \mathbb{R}$. Before proving these results, we claim that

$$\text{for any rational } p < q, \text{ we have } a^p < a^q. \tag{4.14}$$

Indeed, $1 < a$ and $q - p > 0$, so by our power rules,

$$1 = 1^{q-p} < a^{q-p},$$

which, after multiplication by a^p, gives our claim. We now prove that $f \to \infty$ as $x \to \infty$. To prove this, we note that since $a > 1$, we can write $a = 1 + b$ for some $b > 0$, so by Bernoulli's inequality, for any $n \in \mathbb{N}$,

$$a^n = (1 + b)^n \geq 1 + n\,b > n\,b. \tag{4.15}$$

Now fix $M > 0$. By the Archimedean principle, we can choose $N \in \mathbb{N}$ such that $N\,b > M$, therefore by (4.15) and (4.14),

$$x \in \mathbb{Q} \quad \text{and} \quad x > N \quad \Longrightarrow \quad M < N\,b < a^N < a^x.$$

This proves that $f \to \infty$ as $x \to \infty$. We now show that $f \to 0$ as $x \to -\infty$. Let $\varepsilon > 0$. Then by the Archimedean principle there is an $N \in \mathbb{N}$ such that

$$\frac{1}{b\,\varepsilon} < N \quad \Longrightarrow \quad \frac{1}{N\,b} < \varepsilon.$$

[2] *I protest against the use of infinite magnitude as something completed, which in mathematics is never permissible. Infinity is merely a facon de parler, the real meaning being a limit which certain ratios approach indefinitely near, while others are permitted to increase without restriction. Carl Friedrich Gauss (1777–1855).*

By (4.15) and (4.14) it follows that

$$x \in \mathbb{Q} \quad \text{and} \quad x < -N \quad \Longrightarrow \quad 0 < a^x < a^{-N} = \frac{1}{a^N} < \frac{1}{N\,b} < \varepsilon.$$

This proves that $f \to 0$ as $x \to -\infty$.

Many of the limit theorems in Section 4.1.3 that we have worked out for "regular limits" also hold for left and right-hand limits, limits at infinity, and infinite limits. To avoid repeating these limit theorems in each of our new contexts (which will take up a few pages at the least!), we shall make the following general comment:

> *The sequence criterion, uniqueness of limits, component theorem, algebra of limits, composition of limits, squeeze theorem, and preservation of inequalities have analogous statements for left/right-hand limits, limits at infinity, and infinite (properly divergent) limits.*

Of course, some statements don't hold when we consider infinite limits, for example, we cannot subtract infinities or divide them, nor can we multiply zero and infinity. We encourage the reader to think about these analogous statements and we shall make use of these extended versions without much comment in the sequel.

Example 4.13. For an example of this general comment, suppose that $L = \lim_{y\to\infty} f(y)$. We leave the reader to show that $\lim_{x\to 0^+} 1/x = \infty$. Therefore, according to our extended composition of limits theorem, we have

$$L = \lim_{x\to 0^+} f\left(\frac{1}{x}\right).$$

Similarly, since $\lim_{x\to-\infty} -x = \infty$, again by our extended composition of limits theorem, we have

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

More generally, if g is any function with $\lim_{x\to c} g(x) = \infty$ where c is either a real number, ∞, or $-\infty$, then by our composition of limits theorem,

$$L = \lim_{x\to c} f \circ g.$$

Example 4.14. One last example. Suppose as before that $\lim_{x\to c} g(x) = \infty$ where c is either a real number, ∞, or $-\infty$. Since $\lim_{y\to\infty} 1/y = 0$, by our extended composition of limits theorem, we have

$$\lim_{x\to c} \frac{1}{g(x)} = 0.$$

Exercises 4.2.

1. Using the ε-δ definition of (left/right-hand) limit, prove (a) and (b):

   $$(a)\ \lim_{x\to 0-} \frac{x}{|x|} = -1, \quad (b)\ \lim_{x\to 0+} \frac{x}{|x|} = 1.\ \text{Conclude that}\ \lim_{x\to 0} \frac{x}{|x|}\ \text{does not exist.}$$

2. Using the ε-N definition of limits at infinity, prove that

   $$(a)\ \lim_{x\to\infty} \frac{x^2+x+1}{2x^2-1} = \frac{1}{2}, \quad (b)\ \lim_{x\to\infty} \sqrt{x^2+1} - x = 0, \quad (c)\ \lim_{x\to\infty} \sqrt{x^2+x} - x = \frac{1}{2}.$$

3. Let $p(x) = a_n\,x^n + \cdots + a_1\,x + a_0$ and $q(x) = b_m\,x^m + \cdots + b_1\,x + b_0$ be polynomials with real coefficients and with $a_n \neq 0$ and $b_m \neq 0$.
   (a) Prove the for any natural number k, $\lim_{x\to\infty} 1/x^k = 0$.
   (b) If $n < m$, prove that $\lim_{x\to\infty} p(x)/q(x) = 0$.
   (c) If $n = m$, prove that $\lim_{x\to\infty} p(x)/q(x) = a_n/b_n$.

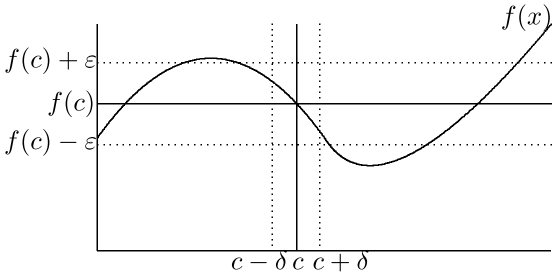


FIGURE 4.3. Visualization of continuity.

(d) If $n > m$, prove that if $a_n > 0$, then $\lim_{x\to\infty} p(x)/q(x) = \infty$, and on the other hand if $a_n < 0$, then $\lim_{x\to\infty} p(x)/q(x) = -\infty$.

4. Let $f, g : D \longrightarrow \mathbb{R}$ with $D \subseteq \mathbb{R}$, $\lim_{x\to\infty} f = \infty$, and $g(x) \neq 0$ for all $x \in D$. Suppose that for some real number L, we have
$$\lim_{x\to\infty} \frac{f}{g} = L.$$
(a) If $L > 0$, prove that $\lim_{x\to\infty} g = +\infty$.
(b) If $L < 0$, prove that $\lim_{x\to\infty} g = -\infty$.
(c) If $L = 0$, can you make any conclusions about $\lim_{x\to\infty} g$?

4.3. Continuity, Thomae's function, and Volterra's theorem

In this section we study the most important functions in all of analysis and topology, continuous functions. We begin by defining what they are and then give examples. Perhaps one of the most fascinating functions you'll ever run across is the modified Dirichlet function or Thomae's function, which has the perplexing and pathological property that it is continuous on the irrational numbers and discontinuous on the rational numbers! We'll see that there is no function opposite to Thomae's, that is, continuous on the rationals and discontinuous on the irrationals; this was proved by Vito Volterra (1860–1940) in 1881. For an interesting account of Thomae's function and its relation to Volterra's theorem, see [**60**].

4.3.1. Continuous functions. We begin by defining continuity at a point. Let $D \subseteq \mathbb{R}^p$. A function $f : D \longrightarrow \mathbb{R}^m$ is **continuous at a point** $c \in D$ if for each $\varepsilon > 0$ there is a $\delta > 0$ such that
$$\boxed{x \in D \quad \text{and} \quad |x - c| < \delta \quad \Longrightarrow \quad |f(x) - f(c)| < \varepsilon.} \tag{4.16}$$
See Figure 4.3 for a picture of what continuity means. The best way to think of a continuous function is that f takes points which are "close" (x and c, which are within δ) to points which are "close" ($f(x)$ and $f(c)$, which are within ε). We can relate this definition to the definition of limit. Suppose that $c \in D$ is a limit point of D. Then comparing (4.16) with the definition of limit, we see that for a limit point c of D such that $c \in D$, we have
$$\boxed{f \text{ is continuous at } c \quad \Longleftrightarrow \quad f(c) = \lim_{x\to c} f.}$$
Technically speaking, when we compare (4.16) to the definition of limit, for a limit we actually require that $0 < |x - c| < \delta$, but in the case that $|x - c| = 0$, that is, $x = c$, we have $|f(x) - f(c)| = |f(c) - f(c)| = 0$, which is automatically less than ε,

so the condition that $0 < |x-c|$ can be dropped. What if $c \in D$ is not a limit point of D? In this case c is called an **isolated point** in D and by definition of (not being a) limit point there is an open ball $B_\delta(c)$ such that $B_\delta(c) \cap D = \{c\}$; that is, the only point of D inside $B_\delta(c)$ is c itself. Hence, with this δ, for any $\varepsilon > 0$, the condition (4.16) is automatically satisfied:

$$x \in D \quad \text{and} \quad |x-c| < \delta \quad \Longrightarrow \quad x = c \quad \Longrightarrow \quad |f(x) - f(c)| = 0 < \varepsilon.$$

Therefore, at isolated points of D, the function f is automatically continuous, and therefore "boring". For this reason, if we want to prove theorems concerning the continuity of $f : D \longrightarrow \mathbb{R}^m$ at a point $c \in D$, we can always assume that c is a limit point of D; in this case, we have all the limit theorems from the last section at our disposal. This is exactly why we spent so much time on learning limits during the last two sections!

If f is continuous at every point in a subset $A \subseteq D$, we say that f is **continuous on** A; in particular, if f is continuous at every point of D, we say that f is **continuous**, or **continuous on** D, to emphasize D:

$$\boxed{f \text{ is continuous} \quad \Longleftrightarrow \quad \text{for all } c \in D,\ f \text{ is continuous at } c.}$$

Example 4.15. Dirichlet's function is discontinuous at every point in $\mathbb{R}$ since in Example 4.7 we already proved that $\lim_{x\to c} D(x)$ does not exist at any $c \in \mathbb{R}$.

Example 4.16. Define $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1^2 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq 0 \\ 0 & (x_1, x_2) = 0. \end{cases}$$

From Example 4.6 we know that $\lim_{x\to 0} f = 0$, so f is continuous at 0.

Example 4.17. If we define $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq 0 \\ 0 & (x_1, x_2) = 0, \end{cases}$$

then we already proved that $\lim_{x\to 0} f$ does not exist in Example 4.8. In particular, f is not continuous at 0.

Example 4.18. From Example 4.10, any polynomial function $p : \mathbb{C} \to \mathbb{C}$ is continuous (that is, continuous at every point $c \in \mathbb{C}$). From Example 4.11, any rational function $p(z)/q(z)$, where p and q are polynomials, is continuous at any point $c \in \mathbb{C}$ such that $q(c) \neq 0$.

4.3.2. Continuity theorems. We now state some theorems on continuity. These theorems follow almost without any work from our limit theorems in the previous section, so we shall omit all the proofs. First we note that the sequence criterion for limits of functions (Theorem 4.2) implies the following theorem, which is worth highlighting!

Theorem 4.10 (**Sequence criterion for continuity**). *A function $f : D \longrightarrow \mathbb{R}^m$ is continuous at $c \in D$ if and only if for every sequence $\{a_n\}$ in D with $c = \lim a_n$, we have $f(c) = \lim f(a_n)$.*

We can write the last equality as $f(\lim a_n) = \lim f(a_n)$ since $c = \lim a_n$. Thus,

$$\boxed{f\left(\lim_{n\to\infty} a_n\right) = \lim_{n\to\infty} f(a_n);}$$

in other words, limits can be "pulled-out" of continuous functions.

Example 4.19. **Question:** Suppose that $f, g : \mathbb{R} \to \mathbb{R}^m$ are continuous and $f(r) = g(r)$ for all rational numbers r; must $f(x) = g(x)$ for all irrational numbers x? The answer is yes, for let c be an irrational number. Then (see Example 3.13 in Section 3.2) there is a sequence of rational numbers $\{r_n\}$ converging to c. Since f and g are both continuous and $f(r_n) = g(r_n)$ for all n, we have

$$f(c) = \lim f(r_n) = \lim g(r_n) = g(c).$$

Note that the answer is false if either f or g were not continuous. For example, with D denoting Dirichlet's function, $D(r) = 1$ for all rational numbers, but $D(x) \neq 1$ for all irrational numbers x. See Problem 2 for a related problem.

The component theorem (Theorem 4.4) implies that a function $f = (f_1, \ldots, f_m)$ is continuous at c if and only if every component f_k is continuous at c.

Theorem 4.11 (**Component criterion for continuity**). *A function is continuous at c if and only if all of its component functions are continuous at c.*

Next, the composition of limits theorem (Theorem 4.6) implies the following theorem.

Theorem 4.12. *Let $f : D \longrightarrow \mathbb{R}^m$ and $g : C \longrightarrow \mathbb{R}^p$ where $D \subseteq \mathbb{R}^p$ and $C \subseteq \mathbb{R}^q$ and suppose that $g(C) \subseteq D$ so that $f \circ g : C \longrightarrow \mathbb{R}^m$ is defined. If g is continuous at c and f is continuous at $g(c)$, then the composite function $f \circ g$ is continuous at c.*

In simple language: *The composition of continuous functions is continuous.* Finally, our algebra of limits theorem (Theorem 4.5) implies the following.

Theorem 4.13. *If $f, g : D \longrightarrow \mathbb{R}^m$ are both continuous at c, then*

(1) $|f|$ and $af + bg$ are continuous at c, for any real a, b.
If f and g take values in $\mathbb{C}$, then
(2) fg and (provided $g(c) \neq 0$) f/g are continuous at c.

In simple language: Linear combinations of $\mathbb{R}^m$-valued continuous functions are continuous. Products, norms, and quotients of real or complex-valued continuous functions are continuous (provided that the denominator functions are not zero). Finally, the left and right-hand limit theorem (Theorem 4.9) implies

Theorem 4.14. *Let $f : D \longrightarrow \mathbb{R}^m$ with $D \subseteq \mathbb{R}$ and let $c \in D$ be a limit point of the sets $D \cap (-\infty, c)$ and $D \cap (c, \infty)$. Then f is continuous at c if and only if $f(c) = f(c+) = f(c-)$.*

If only one of $f(c-)$ or $f(c+)$ makes sense, then f is continuous at c if and only if $f(c) = f(c-)$ or $f(c) = f(c+)$, whichever makes sense.

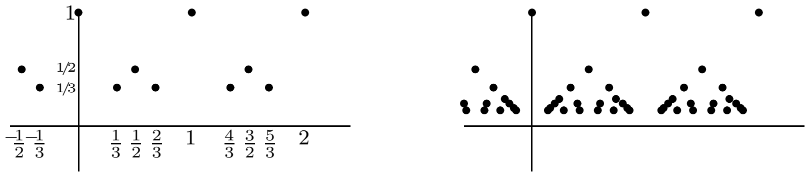


FIGURE 4.4. The left-hand side shows plots of $T(p/q)$ for q at most 3 and the right shows plots of $T(p/q)$ for q at most 7.

4.3.3. Thomae's function and Volterra's theorem. We now define a fascinating function sometimes called Thomae's function [**17**, p. 123] after Johannes Karl Thomae (1840–1921) who published it in 1875 or the (modified) Dirichlet function [**238**], which has the perplexing property that it is continuous at every irrational number and discontinuous at every rational number. (See Problem 7 in Exercises 4.5 for a generalization.)

We define **Thomae's function**, aka (also known as) the **modified Dirichlet function**, $T : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$T(x) = \begin{cases} 1/q & \text{if } x \in \mathbb{Q} \text{ and } x = p/q \text{ in lowest terms and } q > 0, \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Here, we interpret 0 as 0/1 in lowest terms, so $T(0) = 1/1 = 1$. See Figure 4.4 for a graph of this "pathological function." To see that T is discontinuous on rational numbers, let $c \in \mathbb{Q}$ and let $\{a_n\}$ be a sequence of irrational numbers converging to c, e.g. $a_n = c + \sqrt{2}/n$ works (or see Example 3.13 in Section 3.2). Then $\lim T(a_n) = \lim 0 = 0$, while $T(c) > 0$, hence T is discontinuous at c.

To see that T is continuous at each irrational number, c be an irrational number and let $\varepsilon > 0$. Consider the case when $c > 0$ (the case when $c < 0$ is analogous) and choose $m \in \mathbb{N}$ with $c < m$. Let $0 < x \le m$ and let's consider the inequality

$$|T(x) - T(c)| = |T(x) - 0| = T(x) < \varepsilon. \tag{4.17}$$

If x is irrational, then $T(x) = 0 < \varepsilon$ holds. If $x = p/q$ is in lowest terms and $q > 0$, then $T(x) = 1/q < \varepsilon$ holds if and only if $q > 1/\varepsilon$. Set $n := \lfloor 1/\varepsilon \rfloor$, the greatest integer $\le 1/\varepsilon$ (see Section 2.7.3 if you need a reminder of the greatest integer function); then $T(x) = 1/q < \varepsilon$ holds if and only if $q > n$. Summarizing: For $0 < x \le m$, the inequality (4.17) *always holds* unless $x = p/q$ is rational in lowest terms with $0 < q \le n$; that is, when x is in the set

$$\left\{ r \in \mathbb{Q} \; ; \; r = \frac{p}{q} \text{ is in lowest terms}\,, \; 0 < p \le m\,n\,, \; 0 < q \le n \right\}. \tag{4.18}$$

The requirement $0 < p \le m\,n$ is needed so that $0 < p/q \le m$. Now, there are only a finite number of rationals in the set (4.18). In particular, we can choose $\delta > 0$ such that the interval $(c - \delta, c + \delta)$ is contained in $(0, m)$ and contains none of the rational numbers in (4.18). Therefore,

$$x \in (c - \delta, c + \delta) \quad \Longrightarrow \quad |T(x) - T(c)| < \varepsilon,$$

which proves that T is continuous at c. Thus, $T(x)$ is discontinuous at every *rational* number and continuous at every *irrational* number. The inquisitive student might ask if there is a function opposite to Thomae's function, that is,

> Is there a function which is continuous at every *rational* point
> and discontinuous at every *irrational* point?

The answer is "No." There are many ways to prove this; one can answer this question using the Baire category theorem (cf. [**1**, p. 128]), but we shall answer this question using "compactness" arguments originating with Vito Volterra's (1860–1940) first publication in 1881 (before he was twenty!)[**3**]. To state his theorem we need some terminology.

Let $D \subseteq \mathbb{R}^p$. A subset $A \subseteq D$ is said to be **dense** in D if for each point $c \in D$, any open ball centered at c intersects A, that is, for all $r > 0$, $B_r(c) \cap A$ is not empty, or more explicitly, for all $r > 0$, there is a point $x \in A$ such that $|x - c| < r$. This condition is equivalent to the statement that every point in D is either in A or a limit point of A. For $p = 1$, $A \subseteq D$ is dense if for all $c \in D$ and $r > 0$,

$$(c - r, c + r) \cap A \neq \varnothing.$$

Example 4.20. $\mathbb{Q}$ is dense in $\mathbb{R}$, because (by the density of the (ir)rationals in $\mathbb{R}$ — Theorem 2.37) for any $c \in \mathbb{R}$ and $r > 0$, $(c - r, c + r) \cap \mathbb{Q}$ is never empty. Similarly, the set $\mathbb{Q}^c$, the irrational numbers, is also dense in $\mathbb{R}$.

Let $f : D \longrightarrow \mathbb{R}^m$ with $D \subseteq \mathbb{R}^p$ and let $C_f \subseteq D$ denote the set of points in D at which f is continuous. Explicitly,

$$C_f := \{c \in D\,;\, f \text{ is continuous at } c\}\,.$$

The function f is said to be **pointwise discontinuous** if C_f is dense in D.

Theorem 4.15 (**Volterra's theorem**). *On any nonempty open interval, any two pointwise discontinuous functions have a point of continuity in common.*

Proof. Let f and g be pointwise discontinuous functions on an open interval I. We prove our theorem in three steps.

Step 1: A closed interval $[\alpha, \beta]$ is said to be **nontrivial** if $\alpha < \beta$. We first prove that given any $\varepsilon > 0$ and nonempty open interval $(a, b) \subseteq I$, there is a nontrivial closed interval $J \subseteq (a, b)$ such that for all $x, y \in J$,

$$|f(x) - f(y)| < \varepsilon \quad \text{and} \quad |g(x) - g(y)| < \varepsilon.$$

Indeed, since the continuity points of f are dense in I, there is a point $c \in (a, b)$ at which f is continuous, so for some $\delta > 0$, $x \in I \cap (c - \delta, c + \delta)$ implies that $|f(x) - f(c)| < \varepsilon/2$. Choosing $\delta > 0$ smaller if necessary, we may assume that $J' = [c - \delta, c + \delta] \subseteq (a, b)$. Then for any $x, y \in J'$, we have

$$|f(x) - f(y)| = |(f(x) - f(c)) + (f(c) - f(y))| \leq |f(x) - f(c)| + |f(c) - f(y)| < \varepsilon.$$

Using the same argument for g, but with $(c - \delta, c + \delta)$ in place of (a, b) shows that there is a nontrivial closed interval $J \subseteq J'$ such that $x, y \in J$ implies that $|g(x) - g(y)| < \varepsilon$. Since $J \subseteq J'$, the function f automatically satisfies $|f(x) - f(y)| < \varepsilon$ for $x, y \in J$. This completes the proof of **Step 1**.

Step 2: With $\varepsilon = 1$ and $(a, b) = I$ in **Step 1**, there is a nontrivial closed interval $[a_1, b_1] \subseteq I$ such that $x, y \in [a_1, b_1]$ implies that

$$|f(x) - f(y)| < 1 \quad \text{and} \quad |g(x) - g(y)| < 1.$$

Now with $\varepsilon = 1/2$ and $(a, b) = (a_1, b_1)$ in **Step 1**, there is a nontrivial closed interval $[a_2, b_2] \subseteq (a_1, b_1)$ such that $x, y \in [a_2, b_2]$ implies that

$$|f(x) - f(y)| < \frac{1}{2} \quad \text{and} \quad |g(x) - g(y)| < \frac{1}{2}.$$

Continuing by induction, we construct a sequence of nontrivial closed intervals $\{[a_n, b_n]\}$ such that $[a_{n+1}, b_{n+1}] \subseteq (a_n, b_n)$ for each n and $x, y \in [a_n, b_n]$ implies that

$$|f(x) - f(y)| < \frac{1}{n} \quad \text{and} \quad |g(x) - g(y)| < \frac{1}{n}. \tag{4.19}$$

By the nested intervals theorem there is a point c contained in every $[a_n, b_n]$.

Step 3: We now complete the proof. We claim that both f and g are continuous at c. To prove continuity, let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ with $1/n < \varepsilon$. Since $[a_{n+1}, b_{n+1}] \subseteq (a_n, b_n)$, we have $c \in (a_n, b_n)$ so we can choose $\delta > 0$ such that $(c - \delta, c + \delta) \subseteq (a_n, b_n)$. With this choice of $\delta > 0$, in view of (4.19) and the fact that $1/n < \varepsilon$, we obtain

$$|x - c| < \delta \quad \Longrightarrow \quad |f(x) - f(c)| < \varepsilon \quad \text{and} \quad |g(x) - g(c)| < \varepsilon.$$

Thus, f and g are continuous at c and our proof is complete. □

Thus, there cannot be a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that is continuous at every rational point and discontinuous at every irrational point. Indeed, if so, then f would be pointwise discontinuous (because $\mathbb{Q}$ is dense in $\mathbb{R}$) and the function f and Thomae's function wouldn't have any continuity points in common, contradicting Volterra's theorem.

Exercises 4.3.

1. Recall that $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. Determine the set of continuity points for the following functions:

$$(a)\ f(x) = \lfloor x \rfloor, \quad (b)\ g(x) = x\lfloor x \rfloor, \quad (c)\ h(x) = \lfloor 1/x \rfloor,$$

where the domains are $\mathbb{R}$, $\mathbb{R}$, and $(0, \infty)$, respectively. Are the functions continuous on the domains $(-1, 1)$, $(-1, 1)$, and $(1, \infty)$, respectively?

2. In this problem we deal with zero sets of functions. Let $f : D \longrightarrow \mathbb{R}^m$ with $D \subseteq \mathbb{R}^p$. The **zero set** of f is the set $Z(f) := \{x \in D\,;\, f(x) = 0\}$.
   (a) Let $f : D \longrightarrow \mathbb{R}^m$ be continuous and let $c \in D$ be a limit point of $Z(f)$. Prove that $f(c) = 0$.
   (b) Let $f : D \longrightarrow \mathbb{R}^m$ be continuous and suppose that $Z(f)$ is dense in D. Prove that f is the zero function, that is, $f(x) = 0$ for all $x \in D$.
   (c) Using (b), prove that if $f, g : D \longrightarrow \mathbb{R}^m$ are continuous and $f(x) = g(x)$ on a dense subset of D, then $f = g$, that is, $f(x) = g(x)$ for all $x \in D$.

3. In this problem we look at additive functions. Let $f : \mathbb{R}^p \longrightarrow \mathbb{R}^m$ be **additive** in the sense that $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}^p$. Prove:
   (a) Prove that $f(0) = 0$ and that $f(x - y) = f(x) - f(y)$ for all $x, y \in \mathbb{R}^p$.
   (b) If f is continuous at some $x_0 \in \mathbb{R}^p$, then f is continuous on all of $\mathbb{R}^p$.
   (c) Assume now that $p = 1$ so that $f : \mathbb{R} \longrightarrow \mathbb{R}^m$ is additive (no continuity assumptions at this point). Prove that $f(r) = f(1)\, r$ for all $r \in \mathbb{Q}$.
   (d) (Cf. [**251**]) If $f : \mathbb{R} \longrightarrow \mathbb{R}^m$ is continuous, prove that $f(x) = f(1)\, x$ for all $x \in \mathbb{R}$.

4. Let $f : \mathbb{R}^p \longrightarrow \mathbb{C}$ be **multiplicative** in the sense that $f(x + y) = f(x)\, f(y)$ for all $x, y \in \mathbb{R}^p$. Assume that f is not the zero function.
   (a) Prove that $f(x) \neq 0$ for all $x \in \mathbb{R}^p$.
   (b) Prove that $f(0) = 1$ and prove that $f(-x) = 1/f(x)$.
   (c) Prove that if f is continuous at some point x_0, then f is continuous on all of $\mathbb{R}^p$.
   When $p = 1$ and f is real-valued and continuous, in Problem 11 in Exercises 4.6 we show that f is given by an "exponential function".

5. Let $f : I \longrightarrow \mathbb{R}$ be a continuous function on a closed and bounded interval I. Suppose there is a $0 < r < 1$ having the property that for each $x \in I$ there is a point $y \in I$ with

$|f(y)| \le r|f(x)|$. Prove that f must have a root, that is, there is a point $c \in I$ such that $f(c) = 0$.

6. Consider the following function related Thomae's function:
$$t(x) := \begin{cases} q & \text{if } x = p/q \text{ in lowest terms and } q > 0, \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$
Prove that $t : \mathbb{R} \longrightarrow \mathbb{R}$ is discontinuous at every point in $\mathbb{R}$.
7. Here are some fascinating questions related to Volterra's theorem.
   (a) Are there functions $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ that don't have any continuity points in common, one that is pointwise discontinuous and the other one that is not (but is continuous at least at one point)? Give an example or prove there are no such functions.
   (b) Is there a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that maps rationals to irrationals? Give an example or prove there is no such function.
   (c) Is there a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that maps rationals to irrationals *and* irrationals to rationals? Suggestion: Suppose there is such a function and consider the function $T \circ f$ where T is Thomae's function.

4.4. Compactness, connectedness, and continuous functions

Consider the continuous function $f(x) = 1/x$ with domain $D = \mathbb{R} \setminus \{0\}$, the real line with a "hole." By drawing a graph of this function, it will be apparent that f has the following "bad" properties:[3] f is not bounded on D, in particular f does not attain a maximum or minimum value on D, and that although the range of f contains both positive and negative values, f never takes on the intermediate value of 0. In this section we prove that the "bad" nonboundedness property is absent when the domain is a closed and bounded interval and the "bad" nonintermediate value property is absent when the domain is any interval. We shall prove these results in two rather distinct viewpoints using:

(I) A somewhat concrete *analytical* approach that only uses concepts we've covered in previous sections.

(II) A somewhat abstract *topological* approach based on the topological lemmas presented in Section 4.4.1.

If you're only interested in the easier analytical approach, skip Section 4.4.1 and also skip the **Proof II**'s in Theorems 4.19, 4.20, and 4.22. For an interesting and different approach using the concept of "tagged partitions," see Gordon [**84**].

4.4.1. Some fundamental topological lemmas. Let $A \subseteq \mathbb{R}$. A collection $\mathscr{U}$ of subsets of $\mathbb{R}$ is called a **cover** of A (or **covers** A) if the union of all the sets in $\mathscr{U}$ contains A. Explicitly, $\mathscr{U} = \{\mathcal{U}_\alpha\}$ covers A if $A \subseteq \bigcup_\alpha \mathcal{U}_\alpha$. We are mostly interested in coverings by open intervals, that is, where each $\mathcal{U}_\alpha$ is an open interval.

Example 4.21. $(0,1)$ is covered by $\mathscr{U} = \{U_n = (1/n, 1)\}$ because $(0,1) \subseteq \bigcup_{n=1}^\infty (1/n, 1)$. (The diligent student will supply the details!)

Example 4.22. $[0,1]$ is covered by $\mathscr{V} = \{V_n = (-1/n, 1 + 1/n)\}$ because $[0,1] \subseteq \bigcup_{n=1}^\infty (-1/n, 1 + 1/n)$.

It's interesting to notice that $\mathscr{U}$ does not have a **finite subcover**, that is, there are not finitely many elements of $\mathscr{U}$ that will still cover $(0,1)$. To see this, let $\{U_{n_1}, \ldots, U_{n_k}\}$ be a finite subcollection of elements of $\mathscr{U}$. By relabelling, we may assume that $n_1 < n_2 < \cdots < n_k$. Since n_k is the largest of these k numbers,

[3] "Bad" from one angle, from another angle, these properties can be viewed as interesting.

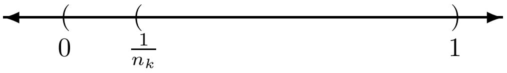


FIGURE 4.5. The finite subcover $\{U_{n_1}, \ldots, U_{n_k}\}$ does not cover $[0, 1)$.

we have $\bigcup_{j=1}^{k} U_{n_j} = (\frac{1}{n_k}, 1)$, which does not cover $(0, 1)$ because there is a "gap" between 0 and $1/n_k$ as seen in Figure 4.5. On the other hand, $\mathscr{V}$ does have a finite subcover, that is, there are finitely many elements of $\mathscr{V}$ that will cover $[0, 1]$. Indeed, $[0, 1]$ is covered by the single element V_1 of $\mathscr{V}$ because $[0, 1] \subseteq (-1, 1 + 1)$. This is in fact a general phenomenon for closed and bounded intervals.

LEMMA 4.16 (**Compactness lemma**). *Every cover of a closed and bounded interval by open intervals has a finite subcover.*

PROOF. Let $\mathscr{U}$ be a cover of $[a, b]$ by open intervals. We must show that there are finitely many elements of $\mathscr{U}$ that still cover $[a, b]$. Let A be the set of all numbers x in $[a, b]$ such that the interval $[a, x]$ is contained in a union of finitely many sets in $\mathscr{U}$. Since $[a, a]$ is the single point a, this interval is contained in a single set in $\mathscr{U}$, so A is not empty. Being a nonempty subset of $\mathbb{R}$ bounded above by b, A has a supremum, say $\xi \leq b$. Since ξ belongs to the interval $[a, b]$ and $\mathscr{U}$ covers $[a, b]$, ξ belongs to some open interval (c, d) in the collection $\mathscr{U}$. Choose any real number η with $c < \eta < \xi$. Then η is less than the supremum of A, so $[a, \eta]$ is covered by finitely many sets in $\mathscr{U}$, say $[a, \eta] \subseteq U_1 \cup \cdots \cup U_k$. Adding $U_{k+1} := (c, d)$ to this collection, it follows that for any real number x with $c < x < d$, the interval $[a, x]$ is covered by the finitely many sets $U_1, \ldots, U_{k+1}$ in $\mathscr{U}$. In particular, since $c < \xi < d$, for any x with $\xi \leq x < d$, the interval $[a, x]$ is covered by finitely many sets in $\mathscr{U}$, so unless $\xi = b$, the set A would contain a number greater than ξ. Hence, $b = \xi$ and $[a, b]$ can be covered by finitely many sets in $\mathscr{U}$. □

Because closed and bounded intervals have this *finite* subcover property, and therefore behave somewhat like finite sets (which are "compact" — take up little space), we call such intervals **compact**. We now move to open sets. An **open set** in $\mathbb{R}$ is simply a union of open intervals; explicitly, $A \subseteq \mathbb{R}$ is open means that $A = \bigcup_\alpha U_\alpha$ where each U_α is an open interval.

Example 4.23. $\mathbb{R} = (-\infty, \infty)$ is an open interval, so $\mathbb{R}$ is open.

Example 4.24. Any open interval (a, b) is open because it's a union consisting of just itself. In particular, if $b \leq a$, we have $(a, b) = \varnothing$, so $\varnothing$ an open set.

Example 4.25. Another example is $\mathbb{R} \setminus \mathbb{Z}$ because $\mathbb{R} \setminus \mathbb{Z} = \bigcup_{n \in \mathbb{Z}} (n, n + 1)$.

A set $A \subseteq \mathbb{R}$ is **disconnected** if there are open sets $\mathcal{U}$ and $\mathcal{V}$ such that $A \cap \mathcal{U}$ and $A \cap \mathcal{V}$ are nonempty, disjoint, and have union A. To have union A, we mean $A = (A \cap \mathcal{U}) \cup (A \cap \mathcal{V})$, which is actually equivalent to saying that

$$A \subseteq \mathcal{U} \cup \mathcal{V}.$$

A set $A \subseteq \mathbb{R}$ is **connected** if it's not disconnected.

Example 4.26. $A = (-1, 0) \cup (0, 1)$ is disconnected because $A = \mathcal{U} \cup \mathcal{V}$ where $\mathcal{U} = (-1, 0)$ and $\mathcal{V} = (0, 1)$ are open, and $A \cap \mathcal{U} = (-1, 0)$ and $A \cap \mathcal{V} = (0, 1)$ are nonempty, disjoint, and union to A.

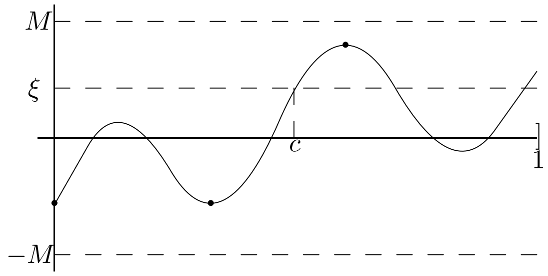


FIGURE 4.6. Illustrations of the boundedness, max/min value, and intermediate value theorems for a function f on $[0, 1]$.

Intuitively, intervals should always be connected. This is in fact the case.

LEMMA 4.17 (**Connectedness lemma**). *Intervals (open, closed, bounded, unbounded, etc.) are connected.*

PROOF. Let I be an interval and suppose, for sake of contradiction, that it is disconnected. Then there are open sets $\mathcal{U}$ and $\mathcal{V}$ such that $I \cap \mathcal{U}$ and $I \cap \mathcal{V}$ are disjoint, nonempty, and have union I. Let $a, b \in I$ with $a \in \mathcal{U}$ and $b \in \mathcal{V}$. By symmetry we may assume that $a < b$. Then $[a, b] \subseteq I$, so $[a, b] \cap \mathcal{U}$ and $[a, b] \cap \mathcal{V}$ are disjoint, nonempty, and have union $[a, b]$. Thus, $[a, b]$ is disconnected so we may as well assume that $I = [a, b]$ in the first place and derive a contradiction from this. Define

$$c := \sup \big(I \cap \mathcal{U}\big).$$

This number exists because $I \cap \mathcal{U}$ contains a and is bounded above by b. In particular, $c \in I$. Since $I \subseteq \mathcal{U} \cup \mathcal{V}$, the point c must belong to either $\mathcal{U}$ or $\mathcal{V}$. We shall derive a contradiction in either situation. Suppose that $c \in \mathcal{U}$. Since $b \in I \cap \mathcal{V}$ and $I \cap \mathcal{U}$ and $I \cap \mathcal{V}$ are disjoint, it follows that $c \neq b$, so $c < b$. Now $\mathcal{U}$ is open, so it's a union of open intervals, therefore $c \in (\alpha, \beta)$ for some open interval (α, β) making up $\mathcal{U}$. This implies that $I \cap \mathcal{U}$ contains points between c and β. However, this is impossible because c is an upper bound for $I \cap \mathcal{U}$. So, suppose that $c \in \mathcal{V}$. Since $a \in I \cap \mathcal{U}$ and $I \cap \mathcal{U}$ and $I \cap \mathcal{V}$ are disjoint, it follows that $c \neq a$, so $a < c$. Now $c \in (\alpha', \beta')$ for some open interval (α', β') making up $\mathcal{V}$. Since $\mathcal{U}$ and $\mathcal{V}$ are disjoint and c is an upper bound for $I \cap \mathcal{U}$, there are points between α' and c that are also upper bounds for $I \cap \mathcal{U}$. This too is impossible since c is the least upper bound. □

4.4.2. The boundedness theorem. The geometric content of the boundedness theorem is that the graph of a continuous function f on a closed and bounded interval lies between two horizontal lines, that is, there is a constant M such that $|f(x)| \leq M$ for all x in the interval. Therefore, the graph does not extend infinitely up or down; see Figure 4.6 (the dots and the point c in the figure have to do with the max/min value and intermediate value theorems). The function $f(x) = 1/x$ on $(0, 1]$ or $f(x) = x$ on any unbounded interval shows that the boundedness theorem does not hold when the interval is not closed and bounded. Before proving the boundedness theorem, we need the following lemma.

LEMMA 4.18 (**Inequality lemma**). *Let $f : I \longrightarrow \mathbb{R}$ be a continuous map on an interval I, let $c \in I$, and suppose that $|f(c)| < d$ where $d \in \mathbb{R}$. Then there is an open interval I_c containing c such that for all $x \in I$ with $x \in I_c$, we have $|f(x)| < d$.*

Proof. Let $\varepsilon = d - |f(c)|$ and, using the definition of continuity, choose $\delta > 0$ such that $x \in I$ and $|x - c| < \delta \implies |f(x) - f(c)| < \varepsilon$. Let $I_c = (c - \delta, c + \delta)$. Then given $x \in I$ with $x \in I_c$, we have $|x - c| < \delta$, so

$$|f(x)| = |f(x) - f(c)| + |f(c)| < \varepsilon + |f(c)| = \big(d - |f(c)|\big) + |f(c)| = d,$$

which proves our claim. □

An analogous proof shows that if $a < f(c) < b$, then there is an open interval I_c containing c such that for all $x \in I$ with $x \in I_c$, we have $a < f(x) < b$. Yet another analogous proof shows that if $f : D \longrightarrow \mathbb{R}^m$ with $D \subseteq \mathbb{R}^p$ is a continuous map and $|f(c)| < d$, then there is an open ball B containing c such that for all $x \in D$ with $x \in B$, we have $|f(x)| < d$. We'll leave these generalizations to the interested reader.

Theorem 4.19 (**Boundedness theorem**). *A continuous real-valued function on a closed and bounded interval is bounded.*

Proof. Let f be a continuous function on a closed and bounded interval I.

Proof I: Assume that f is unbounded; we shall prove that f is not continuous. Since f is unbounded, for each natural number n there is a point x_n in I such that $|f(x_n)| \geq n$. By the Bolzano-Weierstrass theorem, the sequence $\{x_n\}$ has a convergent subsequence, say $\{x'_n\}$ that converges to some c in I. By the way the numbers x_n were chosen, it follows that $|f(x'_n)| \to \infty$, which shows that $f(x'_n) \not\to f(c)$, for if $f(x'_n) \to f(c)$, then we would have $|f(c)| = \lim |f(x'_n)| = \infty$, an impossibility because $f(c)$ is a real number. Thus, f is not continuous at c.

Proof II: Given any arbitrary point c in I, we have $|f(c)| < |f(c)| + 1$, so by our inequality lemma there is an open interval I_c containing c such that for each $x \in I_c$, $|f(x)| < |f(c)| + 1$. The collection of all such open intervals $\mathscr{U} = \{I_c \,;\, c \in I\}$ covers I, so by the compactness lemma, there are finitely many open intervals in $\mathscr{U}$ that cover I, say $I_{c_1}, \ldots, I_{c_n}$. Let M be the largest of the values $|f(c_1)| + 1, \ldots, |f(c_n)| + 1$. We claim that f is bounded by M on all of I. Indeed, given $x \in I$, since $I_{c_1}, \ldots, I_{c_n}$ cover I, there is an interval I_{c_k} containing x. Then,

$$|f(x)| < |f(c_k)| + 1 \leq M.$$

Thus, f is bounded. □

4.4.3. The max/min value theorem. The geometric content of our second theorem is that the graph of a continuous function on a closed and bounded interval must have highest and lowest points. The dots in Figure 4.6 show such extreme points; note that there are two lowest points in the figure. The simple example $f(x) = x$ on $(0, 1)$ shows that the max/min theorem does not hold when the interval is not closed and bounded.

Theorem 4.20 (**Max/min value theorem**). *A continuous real-valued function on a closed and bounded interval achieves its maximum and minimum values. That is, if $f : I \longrightarrow \mathbb{R}$ is a continuous function on a closed and bounded interval I, then for some values c and d in the interval I, we have*

$$f(c) \leq f(x) \leq f(d) \quad \textit{for all } x \textit{ in } I.$$

Proof. Define

$$M := \sup\{f(x) \,;\, x \in I\}.$$

This number is finite by the boundedness theorem. We shall prove that there is a number d in $[a,b]$ such that $f(d) = M$. This proves that f achieves its maximum; a related proof shows that f achieves its minimum.

Proof I: By definition of supremum, for each natural number n, there exists an x_n in I such that

$$M - \frac{1}{n} < f(x_n) \leq M, \tag{4.20}$$

for otherwise, the value $M - 1/n$ would be a smaller upper bound for $\{f(x)\,;\, x \in I\}$. By the Bolzano-Weierstrass theorem, the sequence $\{x_n\}$ has a convergent subsequence $\{x'_n\}$; let's say that $x'_n \to d$ where d is in $[a,b]$. By continuity, we have $f(x'_n) \to f(d)$. On the other hand, by (4.20) and the squeeze theorem, we have $f(x_n) \to M$, so $f(x'_n) \to M$ as well. By uniqueness of limits, $f(d) = M$.

Proof II: Assume, for sake of contradiction, that $f(x) < M$ for all x in I. Let c be any point in I. Since $f(c) < M$ by assumption, we can choose $\varepsilon_c > 0$ such that $f(c) + \varepsilon_c < M$, so by our inequality lemma there is an open interval I_c containing c such that for all $x \in I_c$, $|f(x)| < M - \varepsilon_c$. The collection $\mathscr{U} = \{I_c\,;\, c \in I\}$ covers I, so by the compactness lemma, there are finitely many open intervals in $\mathscr{U}$ that cover I, say $I_{c_1}, \ldots, I_{c_n}$. Let m be the largest of the finitely many values $\varepsilon_{c_k} + |f(c_k)|$, $k = 1, \ldots, n$. Then $m < M$ and given any $x \in I$, since $I_{c_1}, \ldots, I_{c_n}$ cover I, there is an interval I_{c_k} containing x, which shows that

$$|f(x)| < \varepsilon_{c_k} + |f(c_k)| \leq m < M.$$

This implies that M cannot be the supremum of f over I, since m is a smaller upper bound for f. This gives a contradiction to the definition of M. □

4.4.4. The intermediate value theorem. A real-valued function f on an interval I is said to have the **intermediate value property** if it attains all its intermediate values in the sense that if $a < b$ both belong to I, then given any real number ξ between $f(a)$ and $f(b)$, there is a c in $[a,b]$ such that $f(c) = \xi$. By "between" we mean that either $f(a) \leq \xi \leq f(b)$ or $f(b) \leq \xi \leq f(a)$. Geometrically, this means that the graph of f can be draw without "jumps," that is, without ever lifting up the pencil. We shall prove the intermediate value theorem, which states that any continuous function on an interval has the intermediate value property. See the previous Figure 4.6 for an example where we take, for instance, $a = 0$ and $b = 1$; note for this example that the point c need not be unique (there is another c' such that $f(c') = \xi$). The function in the introduction to this section shows that the intermediate value theorem fails when the domain is not an interval.

Before proving the intermediate value theorem we first think a little about intervals. Note that if I is an interval, bounded or unbounded, open, closed, etc, then given any points $a, b \in I$ with $a < b$, it follows that every point c between a and b is also in I. The converse statement: "if $A \subseteq \mathbb{R}$ is such that given any points $a, b \in A$ with $a < b$, every point c between a and b is also in A, then A is an interval" is "obviously" true. We shall leave its proof to the interested reader.

Lemma 4.21. *A set A in $\mathbb{R}$ is an interval if and only if given any points $a < b$ in A, we have $[a,b] \subseteq A$. Stated another way, A is an interval if and only if given any two points a, b in A with $a < b$, all points between a and b also lie in A.*

(In fact, some mathematicians might even take this lemma as the *definition* of interval.) We are now ready to prove our third important theorem in this section.

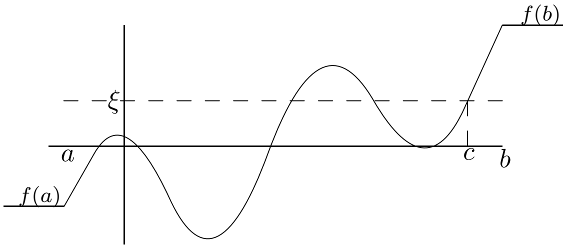


FIGURE 4.7. Proof of the intermediate value property.

THEOREM 4.22 (**Intermediate value theorem**). *A real-valued continuous function on any interval (bounded, unbounded, open, closed, ...) has the intermediate value property. Moreover, the range of the function is also an interval.*

PROOF. Let f be a real-valued continuous function on an interval I and let ξ be between $f(a)$ and $f(b)$ where $a < b$ and $a, b \in I$. We shall prove that there is a c in $[a, b]$ such that $f(c) = \xi$. Assume that $f(a) \leq \xi \leq f(b)$; the reverse inequalities have a related proof. Note that if $\xi = f(a)$, then $c = a$ works or if $\xi = f(b)$, then $c = b$ works, so we may assume that $f(a) < \xi < f(b)$. We now prove that f has the intermediate value property.

Proof I: To prove that f has the IVP we don't care about f outside of $[a, b]$, so let's (re)define f outside of the interval $[a, b]$ such that f is equal to the constant value $f(a)$ on $(-\infty, a)$ and $f(b)$ on (b, ∞). This gives us a continuous function, which we again denote by f, that has domain $\mathbb{R}$ as shown in Figure 4.7.

Define

$$A = \{x \in \mathbb{R}\,;\, f(x) \leq \xi\}.$$

Since $f(a) < \xi$, we see that $a \in A$ so A is not empty and since $\xi < f(b)$, we see that A is bounded above by b. In particular, $c := \sup A$ exists and $a \leq c \leq b$. We shall prove that $f(c) = \xi$, which is "obvious" from Figure 4.7. To prove this rigourously, observe that by definition of *least* upper bound, for any $n \in \mathbb{N}$, there is a point $x_n \in A$ such that $c - \frac{1}{n} < x_n \leq c$. As $n \to \infty$, we have $x_n \to c$, so by continuity, $f(x_n) \to f(c)$. Since $f(x_n) \leq \xi$, because $x_n \in A$, and limits preserve inequalities, we have $f(c) \leq \xi$. On the other hand, by definition of upper bound, for any $n \in \mathbb{N}$, we must have $f(c + \frac{1}{n}) > \xi$. Taking $n \to \infty$ and using that f is continuous and that limits preserve inequalities, we see that $f(c) \geq \xi$. It follows that $f(c) = \xi$.

Proof II: To prove that f has the IVP using topology, suppose that $f(x) \neq \xi$ for any x in I. Let c be any point in I. If $f(c) < \xi$, then by the discussion after our inequality lemma, there is an open interval I_c containing c such that if $x \in I$ with $x \in I_c$, we have $f(x) < \xi$. Similarly, if $\xi < f(c)$, there is an open interval I_c containing c such that if $x \in I$ with $x \in I_c$, we have $\xi < f(x)$. In summary, we have assigned to each point $c \in I$, an open interval I_c that contains c such that either $f(x) < \xi$ or $\xi < f(x)$ for all $x \in I$ with $x \in I_c$. Let $\mathcal{U}$ be the union of all the I_c's where $f(c) < \xi$ and let $\mathcal{V}$ be the union of all the I_c's where $\xi < f(c)$. Then $\mathcal{U}$ and $\mathcal{V}$ are unions of open intervals so are open sets by definition, and $a \in \mathcal{U}$ since $f(a) < \xi$, and $b \in \mathcal{V}$ since $\xi < f(b)$. Notice that $\mathcal{U}$ and $\mathcal{V}$ are disjoint because $\mathcal{U}$ has the property that if $x \in \mathcal{U}$, then $f(x) < \xi$ and $\mathcal{V}$ has the property that if $x \in \mathcal{V}$, then $\xi < f(x)$. Thus, $\mathcal{U}$ and $\mathcal{V}$ are disjoint, nonempty, and $I \subseteq \mathcal{U} \cup \mathcal{V}$. This contradicts the fact that intervals are connected.

We now prove that $f(I)$ is an interval. By our lemma, $f(I)$ is an interval if and only if given any points α, β in $f(I)$ with $\alpha < \beta$, all points between α and β also lie in $f(I)$. Since $\alpha, \beta \in f(I)$, we can write $\alpha = f(x)$ and $\beta = f(y)$. Now let $f(x) < \xi < f(y)$. We need to show that $\xi \in f(I)$. However, according to the intermediate value property, there is a c in I such that $f(c) = \xi$. Thus, ξ is in $f(I)$ and our proof is complete. □

A **root** or **zero** of a function $f : D \longrightarrow \mathbb{R}^m$ (with $D \subseteq \mathbb{R}^p$) is point $c \in D$ such that $f(c) = 0$.

Corollary 4.23. *Let f be a real-valued continuous function on an interval and let $a < b$ be points in the interval such that $f(a)$ and $f(b)$ have opposite signs (that is, $f(a) > 0$ and $f(b) < 0$, or $f(a) < 0$ and $f(b) > 0$). Then there is a number $a < c < b$ such that $f(c) = 0$.*

Proof. Since 0 is between $f(a)$ and $f(b)$, by the intermediate value theorem there is a point c in $[a, b]$ such that $f(c) = 0$; since $f(a)$ and $f(b)$ are nonzero, c must lie strictly between a and b. □

4.4.5. The fundamental theorems of continuous functions in action.

Example 4.27. The intermediate value theorem helps us to solve the following puzzle [**219**, p. 239]. (For another interesting puzzle, see Problem 6.) At 1 o'clock in the afternoon a man starts walking up a mountain, arriving at 10 o'clock in the evening at his hut. At 1 o'clock the next afternoon he walks back down by the exact same route, arriving at 10 o'clock in the evening at the point he started off the day before. Prove that at some time he is at the same place on the mountain on both days. To solve this puzzle, let $f(x)$ and $g(x)$ be the distance of the man from his hut, measured along his route, at time x on day one and two, respectively. Then $f : [1, 10] \longrightarrow \mathbb{R}$ and $g : [1, 10] \longrightarrow \mathbb{R}$ are continuous. We need to show that $f(x) = g(x)$ at some time x. To see this, let $h(x) = f(x) - g(x)$. Then $h(1) = f(1) > 0$ and $h(10) = -g(10) < 0$. The IVP implies there is some point t where $h(t) = 0$. This t is a time that solves our puzzle.[4]

Example 4.28. The intermediate value theorem can be used to prove that any nonnegative real number has a square root. To see this, let $a \geq 0$ and consider the function $f(x) = x^2$. Then f is continuous on $\mathbb{R}$, $f(0) = 0$, and

$$f(a + 1) = (a + 1)^2 = a^2 + 2a + 1 \geq 2a \geq a.$$

Therefore, $f(0) \leq a \leq f(a + 1)$, so by the intermediate value theorem, there is a point $0 \leq c \leq a + 1$ such that $f(c) = a$, that is, $c^2 = a$. This proves that a has a square root. (The uniqueness of c follows from the last power rule in Theorem 2.22.) Of course, considering the function $f(x) = x^n$, we can prove that any nonnegative real number has a unique n-th root.

Example 4.29. Here's an interesting **Question:** Is there a continuous function $f : [0, 1] \longrightarrow \mathbb{R}$ that takes on each value in its range exactly twice? In other words, for each $y \in f([0, 1])$ there are exactly two points $x_1, x_2 \in [0, 1]$ such that $y = f(x_1) = f(x_2)$ — such a function is said to be "two-to-one". The answer is no. (See Problem 8 for generalizations of this example.) To see this, assume, by way

[4] A nonmathematical way to solve this problem is to have the man's "twin" walk down the mountain while the man is walking up. At some moment, the two men will cross.

of contradiction, that there is such a two-to-one function. Let y_0 be the maximum value of f, which exists by the max/min value theorem. Then there are exactly two points $a, b \in [0,1]$, say $0 \le a < b \le 1$, such that $y_0 = f(a) = f(b)$. Note that all other points $x \in [0,1]$ besides a, b must satisfy $f(x) < y_0$. This is because if $x \neq a, b$ yet $f(x) = y_0 = f(a) = f(b)$, then there would be three points $x, a, b \in [0,1]$ taking on the same value contradicting the two-to-one property. We claim that $a = 0$. Indeed, suppose that $0 < a$ and choose any $c \in (a,b)$; then, $0 < a < c < b$. Since $f(0) < y_0$ and $f(c) < y_0$ we can choose a $\xi \in \mathbb{R}$ such that $f(0) < \xi < y_0$ and $f(c) < \xi < y_0$. Therefore,

$$f(0) < \xi < f(a) \quad , \quad f(c) < \xi < f(a) \quad , \quad f(c) < \xi < f(b).$$

By the intermediate value theorem, there are points

$$0 < c_1 < a \quad , \quad a < c_2 < c \quad , \quad c < c_3 < b$$

such that $\xi = f(c_1) = f(c_2) = f(c_3)$. Note that c_1, c_2, c_3 are all distinct and ξ is taken on at least three times by f. This contradicts the two-to-one property, so $a = 0$. Thus, f achieves its maximum at 0. Since $-f$ is also two-to-one, it follows that $-f$ also achieves its maximum at 0, which is the same as saying that f achieves its minimum at 0. However, if $y_0 = f(0)$ is both the maximum and minimum of f, then f must be the constant function $f(x) = y_0$ for all $x \in [0,1]$ contradicting the two-to-one property of f.

Exercises 4.4.

1. Is there a *nonconstant* continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that takes on only rational values (that is, whose range is contained in $\mathbb{Q}$)? What about only irrational values?
2. In this problem we investigate real roots of real-valued *odd* degree polynomials.
   (a) Let $p(x) = x^n + a_{n-1}\,x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial with each a_k real and $n \ge 1$ (not necessarily odd). Prove that there is a real number $a > 0$ such that

$$\frac{1}{2} \le 1 + \frac{a_{n-1}}{x} + \cdots + \frac{a_0}{x^n}, \qquad \text{for all } |x| \ge a. \tag{4.21}$$

   (b) Using (4.21), prove that if n is odd, there is a $c \in [-a, a]$ with $p(c) = 0$.
   (c) **Puzzle:** Does there exist a real number that is one more than its cube?
3. In this problem we investigate real roots of real-valued *even* degree polynomials. Let $p(x) = x^n + a_{n-1}\,x^{n-1} + \cdots + a_1 x + a_0$ with each a_k real and $n \ge 2$ even.
   (a) Let $b > 0$ with $b^n \ge \max\{2a_0, a\}$ where a is given in (4.21). Prove that if $|x| \ge b$, then $p(x) \ge a_0 = p(0)$.
   (b) Prove there is a $c \in \mathbb{R}$ such that for all $x \in \mathbb{R}$, $p(c) \le p(x)$. That is, $p : \mathbb{R} \longrightarrow \mathbb{R}$ achieves a minimum value. Is this statement true for odd-degree polynomials?
   (c) Show that there exists a $d \in \mathbb{R}$ such that the equation $p(x) = \xi$ has a solution $x \in \mathbb{R}$ if and only if $\xi \ge d$. In particular, p has a real root if and only if $d \le 0$.
4. Here are a variety of continuity problems. Let $f : [0,1] \longrightarrow \mathbb{R}$ be continuous.
   (a) If f is one-to-one, prove that f achieves its maximum and minimum values at 0 or 1; that is, the maximum and minimum values of f cannot occur at points in $(0,1)$.
   (b) If $f(0) = f(1)$, prove there are points $a, b \in (0,1)$ with $a \neq b$ such that $f(a) = f(b)$.
   (c) If f is one-to-one and $f(0) < f(1)$, prove that f is strictly increasing; that is, for all $a, b \in [0,1]$ with $a < b$, we have $f(a) < f(b)$.
   (d) If $g : [0,1] \longrightarrow \mathbb{R}$ is continuous such that $f(0) < g(0)$ and $g(1) > f(1)$, prove that there is a point $c \in (0,1)$ such that $f(c) = g(c)$.
5. (**Brouwer's fixed point theorem**) If $f : [a,b] \longrightarrow [a,b]$ is a continuous function, prove that there is a point $c \in [a,b]$ such that $f(c) = c$. This result is a special case of a theorem by Luitzen Egbertus Jan Brouwer (1881–1966). **Puzzle:** You are given a straight wire lying perpendicular to a wall and you bend it into any shape you can

imagine and put it back next to the wall. Is there a point on the bent wire whose distance to the wall is exactly the same as it was originally?

6. (**Antipodal point puzzle**) Prove that there are, at any given moment, antipodal points on the earth's equator that have the same temperature. Here are some steps:
   (i) Let $a > 0$ and let $f : [0, 2a] \longrightarrow \mathbb{R}$ be a continuous function with $f(0) = f(2a)$. Show that there exists a point $\xi \in [0, a]$ such that $f(\xi) = f(\xi + a)$.
   (ii) Using (i) solve our puzzle.
7. Let $f : I \longrightarrow \mathbb{R}$ and $g : I \longrightarrow \mathbb{R}$ be continuous functions on a closed and bounded interval I and suppose that $f(x) < g(x)$ for all x in I
   (a) Prove that there is a constant $\alpha > 0$ such that $f(x) + \alpha < g(x)$ for all $x \in I$.
   (b) Prove that there is a constant $\beta > 1$ such that $\beta f(x) < g(x)$ for all $x \in I$.
   (c) Do properties (a) and (b) hold in case I is bounded but not closed (e.g. $I = (0, 1)$ or $I = (0, 1]$) or unbounded (e.g. $I = \mathbb{R}$ or $I = [1, \infty)$)? In each of these two cases prove (a) and (b) are true, or give counterexamples.
8. (**n-to-one functions**) This problem is a continuation of Example 4.29.
   (a) Define a (necessarily non-continuous) function $f : [0, 1] \longrightarrow \mathbb{R}$ that takes on each value in its range exactly two times.
   (b) Prove that there does not exist a function $f : [0, 1] \longrightarrow \mathbb{R}$ that takes on each value in its range exactly n times, where $n \in \mathbb{N}$ with $n \geq 2$.
   (c) Now what about a function with domain $\mathbb{R}$ instead of $[0, 1]$? Prove that there does not exist a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that takes on each value in its range exactly two times.
   (d) Prove that there does not exist a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that takes on each value in its range exactly n times, where $n \in \mathbb{N}$ is even. If n is odd, there does exist such a function! Draw such a function when $n = 3$ (try to draw a "zig-zag" type function). If you're interested in a formula for a continuous n-to-one function for arbitrary odd n, try to come up with one or see Wenner [**242**].
9. Show that a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ can have at most a countable number of strict maxima. Here, a **strict maximum** is a point c such that $f(x) < f(c)$ for all x sufficiently close to c. Suggestion: At each point c where f has a strict maximum, choose an interval (p, q) containing c where $p, q \in \mathbb{Q}$.

The remaining exercises give alternative proofs of the boundedness, max/min, and intermediate value theorems.

10. (**Boundedness, Proof III**) We shall give another proof of the boundedness theorem as follows. Let f be a real-valued continuous function on a closed interval $[a, b]$. Define
$$A = \{c \in [a, b]\,;\, f \text{ is a bounded on } [a, c]\}.$$
If we prove that $b \in A$, then f is bounded on $[a, b]$, which proves our theorem.
   (i) Show that $a \in A$ and $d := \sup A$ exists where $d \leq b$. We show that $d = b$.
   (ii) Suppose that $d < b$. Show that there is an open interval I containing d such that f is bounded on $[a, b] \cap I$, and moreover, for all points $c \in I$ with $d < c < b$, f is bounded on $[a, c]$. Derive a contradiction.
11. (**Max/min, Proof III**) We give another proof of the max/min value theorem as follows. Let M be the supremum of a real-valued continuous function f on a closed and bounded interval I. Assume that $f(x) < M$ for all x in I and define
$$g(x) = \frac{1}{M - f(x)}.$$
Show that g is continuous on I. However, show that g is actually not bounded on I. Now use the boundedness theorem to arrive at a contradiction.
12. (**Max/min, Proof IV**) Here's a proof of the max/min value theorem that is similar to the proof of the boundedness theorem in Problem 10. For each $c \in [a, b]$ define
$$M_c = \sup\{f(x)\,;\, x \in [a, c]\}.$$

This number is finite by the boundedness theorem and $M := M_b$ is the supremum of f over all of $[a, b]$. We shall prove that there is a number d in $[a, b]$ such that $f(d) = M$. This proves that f achieves its maximum; a related proof shows that f achieves its minimum. Define

$$A = \{c \in [a, b]\, ;\, M_c < M\}.$$

(i) If $a \notin A$, prove that $f(a) = M$ and we are done.

(ii) So, suppose that $a \in A$. Show that $d := \sup A$ exists where $d \leq b$. We claim that $f(d) = M$. By way of contradiction, suppose that $f(d) < M$. Let $\varepsilon > 0$ satisfy $f(d) < M - \varepsilon$ and, by the inequality lemma, choose an open interval I containing d such that for all $x \in [a, b]$ with $x \in I$, $f(x) < M - \varepsilon$. Show that there is an $m < M$ such that for any $c \in [a, b]$ with $c \in I$, $M_c < m$. In the two cases, $d < b$ or $d = b$, derive a contradiction.

13. (**IVP, Proof III**) Here's another proof of the intermediate value property. Let f be a real-valued continuous function on an interval $[a, b]$ and suppose that $f(a) < \xi < f(b)$.

(i) Define

$$A = \{x \in [a, b]\, ;\, f(x) < \xi\}.$$

Show that $c := \sup A$ exists. We shall prove that $f(c) = \xi$. Indeed, either this holds or $f(c) < \xi$ or $f(c) > \xi$.

(ii) If $f(c) < \xi$, derive a contradiction by showing that c is not an upper bound.

(iii) If $f(c) > \xi$, derive a contradiction by showing that c is not the least upper bound.

14. (**IVP, Proof IV**) In this problem we prove the intermediate value theorem using the compactness lemma. Let f be a real-valued continuous function on $[a, b]$ and let $f(a) < \xi < f(b)$. Suppose that $f(x) \neq \xi$ for all x in $[a, b]$.

(i) Let $\mathscr{U}$ be the collection of all the open intervals I_c constructed in Theorem 4.22. This collection covers $[a, b]$, so by the compactness lemma, there are finitely many open intervals in $\mathscr{U}$ that cover I, say $(a_1, b_1), \ldots, (a_n, b_n)$. We may assume that

$$a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n$$

by reordering the a_k's if necessary. Prove that $f(x) < \xi$ for all $x \in (a_1, b_1)$.

(ii) Using induction, prove that $f(x) < \xi$ for all x in (a_k, b_k), $k = 1, \ldots, n$. Derive a contradiction, proving the intermediate value theorem.

15. (**IVP, Proof V**) Finally, we give one last proof of the intermediate value theorem called the "bisection method". Let f be a continuous function on an interval and suppose that $a < b$ and $f(a) < \xi < f(b)$.

(i) Let $a_1 = a$ and $b_1 = b$ and let c_1 be the midpoint of $[a_1, b_1]$ and define the numbers a_2 and b_2 by $a_2 = a_1$ and $b_2 = c_1$ if $\xi \leq f(c_1)$ or $a_2 = c_1$ and $b_2 = b_1$ if $f(c_1) < \xi$. Prove that in either case, we have $f(a_2) < \xi \leq f(b_2)$.

(ii) Using $[a_2, b_2]$ instead of $[a_1, b_1]$ and c_2 the midpoint of $[a_2, b_2]$ instead of c_1, and so on, construct a nested sequence of closed and bounded intervals $[a_n, b_n]$ such that $f(a_n) < \xi \leq f(b_n)$ for each n.

(iii) Using the nested intervals theorem show that the intersection of all $[a_n, b_n]$ is a single point, call it c, and show that $f(c) = \xi$.

16. We prove the connectedness lemma using the notion of "chains". Let $\mathcal{U}$ and $\mathcal{V}$ be open sets and suppose that $[a, b] \cap \mathcal{U}$ and $[a, b] \cap \mathcal{V}$ are disjoint, nonempty, and have union $[a, b]$. We define a **chain** in $\mathcal{U}$ as finitely many intervals $I_1, \ldots, I_n$, where the I_k's are open intervals in the union defining $\mathcal{U}$ (recall that $\mathcal{U}$, being open, is by definition a union of open intervals), such that $a \in I_1$ and $I_k \cap I_{k+1} \neq \varnothing$ for $k = 1, \ldots, n - 1$. Let

$$A = \big\{c \in [a, b]\, ;\, \text{there is a chain } I_1, \ldots, I_n \text{ in } \mathcal{U} \text{ with } c \in I_n.\big\}$$

(i) Show that $a \in A$ and that $c = \sup A$ exists, where $c \in [a, b]$. Then $c \in \mathcal{U}$ or $c \in \mathcal{V}$.

(ii) However, show that $c \notin \mathcal{U}$ by assuming $c \in \mathcal{U}$ and deriving a contradiction.

(iii) However, show that $c \notin \mathcal{V}$ by assuming $c \in \mathcal{V}$ and deriving a contradiction.

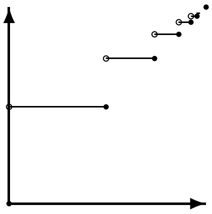


FIGURE 4.8. Zeno's function $Z : [0,1] \longrightarrow \mathbb{R}$.

4.5. Monotone functions and their inverses

In this section we study monotone functions on intervals and their continuity properties. In particular, we prove the following fascinating fact: Any monotone function on an interval (no other assumptions besides monotonicity) is continuous everywhere on the interval except perhaps at countably many points. With the monotonicity assumption dropped, anything can happen, for instance, recall that Dirichlet's function is nowhere continuous.

4.5.1. Continuous and discontinuous monotone functions. Let $I \subseteq \mathbb{R}$ be an interval. A function $f : I \longrightarrow \mathbb{R}$ is said to be **nondecreasing** if $a \le b$ (where $a, b \in I$) implies $f(a) \le f(b)$, **(strictly) increasing** if $a < b$ implies $f(a) < f(b)$, **nonincreasing** if $a \le b$ implies $f(a) \ge f(b)$, and **(strictly) decreasing** if $a < b$ implies $f(a) > f(b)$. The function is **monotone** if it's one of these four types. (Really two types because increasing and decreasing functions are special cases of nondecreasing and nonincreasing functions, respectively.)

Example 4.30. A neat example of a monotone (nondecreasing) function is **Zeno's function** $Z : [0,1] \longrightarrow \mathbb{R}$, named after Zeno of Elea (490 B.C.–425 B.C.):

$$Z(x) = \begin{cases} 0 & x = 0 \\ 1/2 & 0 < x \le 1/2 \\ 1/2 + 1/2^2 = 3/4 & 1/2 < x \le 3/4 \\ 1/2 + 1/2^2 + 1/2^3 = 7/8 & 3/4 < x \le 7/8 \\ \cdots \text{ etc. } \cdots & \cdots \\ 1 & x = 1. \end{cases}$$

See Figure 4.8. This function is called Zeno's function because as described by Aristotle (384 B.C.–322 B.C.), Zeno argued that "there is no motion because that which is moved must arrive at the middle of its course before it arrives at the end" (you can read about this in [**100**]). Zeno's function moves from 0 to 1 via half-way stops. Observe that the left-hand limits of Zeno's function exist at each point of $[0,1]$ except at $x = 0$ where the left-hand limit is not defined, and the right-hand limits exist at each point of $[0,1]$ except at $x = 1$ where the right-hand limit is not defined. Also observe that Zeno's function has discontinuity points exactly at the (countably many) points $x = (2^k - 1)/2^k$ for $k = 0, 1, 2, 3, 4, \ldots$.

It's an amazing fact that Zeno's function is typical: *Every* monotone function on an interval has left and right-hand limits at every point of the interval except at the end points when a left or right-hand limit is not even defined and has at most countably many discontinuities. For simplicity ...

> *To avoid worrying about end points, in this section we only consider monotone functions with domain $\mathbb{R}$. However, every result we prove has an analogous statement for domains that are intervals.*

We repeat, every statement we mention in this section holds for monotone functions on intervals (open, closed, half-open, etc.) as long as we make suitable modifications of these statements at end points.

LEMMA 4.24. *Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be nondecreasing. Then the left and right-hand limits, $f(c-) = \lim_{x\to c-} f(x)$ and $f(c+) = \lim_{x\to c+} f(x)$, exist at every point $c \in \mathbb{R}$. Moreover, the following relations hold:*

$$f(c-) \le f(c) \le f(c+),$$

and if $c < d$, then

$$f(c+) \le f(d-). \tag{4.22}$$

PROOF. Fix $c \in \mathbb{R}$. We first show that $f(c-)$ exists. Since f is nondecreasing, for all $x \le c$, $f(x) \le f(c)$, so the set $\{f(x)\,;\, x < c\}$ is bounded above by $f(c)$. Hence, $b := \sup\{f(x)\,;\, x < c\}$ exists and $b \le f(c)$. Given any $\varepsilon > 0$, by definition of supremum there is a $y < c$ such that $b - \varepsilon < f(y)$. Let $\delta = c - y$. Then $c - \delta < x < c$ implies that $y < x < c$, which implies that

$$\begin{aligned} |b - f(x)| &= b - f(x) \quad \text{(since } f(x) \le b \text{ by definition of supremum)}\\ &\le b - f(y) \quad \text{(since } f(y) \le f(x))\\ &< \varepsilon. \end{aligned}$$

This shows that

$$\lim_{x\to c-} f(x) = b = \sup\{f(x)\,;\, x < c\}.$$

Thus, $f(c-)$ exists and $f(c-) = b \le f(c)$. By considering the set $\{f(x)\,;\, c < x\}$ one can similarly prove that

$$f(c+) = \inf\{f(x)\,;\, c < x\} \ge f(c).$$

Let $a < b$. Then given any c with $a < c < b$, by definition of infimum and supremum, we have

$$f(a+) = \inf\{f(x)\,;\, x < a\} \le f(c) \le \sup\{f(x)\,;\, x < b\} = f(b-).$$

Our proof is now complete. □

Of course, there is a corresponding lemma for nonincreasing functions where the inequalities in this lemma are reversed. As a corollary of the property $f(c-) \le f(c) \le f(c+)$ (for a nondecreasing function) and Theorem 4.14, which states that f is continuous at c if and only if $f(c-) = f(c) = f(c+)$, we see that a nondecreasing function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is discontinuous at a point c if and only if $f(c+) - f(c-)$ is positive. In particular, Figure 4.9 shows that there are three basic types of discontinuities that a (in the picture, nondecreasing) monotone function may have. These discontinuities are jump discontinuities, where a function $f : D \longrightarrow \mathbb{R}^m$ with $D \subseteq \mathbb{R}$ is said to have a **jump discontinuity** at a point $c \in D$ if f is discontinuous at c but both the left and right-hand limits $f(c\pm)$ exist, provided that c is a limit point of $D \cap (c, \infty)$ and $D \cap (-\infty, c)$; the number $f(c+) - f(c-)$ is then called the **jump** of f at c. If c is only a limit point of one of the sets $D\cap(c, \infty)$ and $D\cap(-\infty, c)$ then we require only the corresponding right or left-hand limit to exist. Here is a

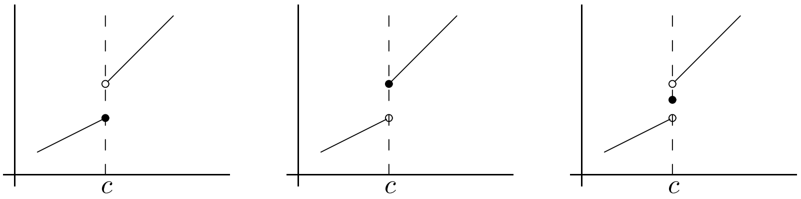


FIGURE 4.9. Monotone functions have only jump discontinuities.

proof that every monotone function has at most countably many discontinuities, each of which being a jump discontinuity; see Problem 2 for another proof.

THEOREM 4.25. *A monotone function on $\mathbb{R}$ has uncountably many points of continuity and at most countably many discontinuities, each discontinuity being a jump discontinuity.*

PROOF. Assume that f is nondecreasing, the case for a nonincreasing function is proved in an analogous manner. We know that f is discontinuous at a point x if and only if $f(x+) - f(x-) > 0$. Given such a discontinuity point, choose a rational number r_x in the interval $(f(x-), f(x+))$. Since f is nondecreasing, given any two such discontinuity points $x < y$, we have (see (4.22)) $f(x+) \leq f(y-)$, so the intervals $(f(x-), f(x+))$ and $(f(y-), f(y+))$ are disjoint. Thus, $r_x \neq r_y$ and to each discontinuity, we have assigned a unique rational number. It follows that the set of all discontinuity points of f is in one-to-one correspondence with a subset of the rationals, and therefore, since a subset of a countable set is countable, the set of all discontinuity points of f is countable. Since $\mathbb{R}$, which is uncountable, is the union of the continuity points of f and the discontinuity points of f, the continuity points of f must be uncountable. □

The following is a very simple and useful characterization of continuous monotone functions on intervals.

THEOREM 4.26. *A monotone function on $\mathbb{R}$ is continuous on $\mathbb{R}$ if and only if its range is an interval.*

PROOF. By the intermediate value theorem, we already know that the range of any (in particular, a monotone) continuous function on $\mathbb{R}$ is an interval. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be monotone and suppose, for concreteness, that f is nondecreasing, the case for a nonincreasing function being similar. It remains to prove that if the range of f is an interval, then f is continuous. We shall prove the contrapositive. So, assume that f is not continuous on I. Then at some point c, we have

$$f(c-) < f(c+).$$

Since f is nondecreasing, this inequality implies that either interval $(f(c-), f(c))$ or $(f(c), f(c+))$, whichever is nonempty, is not contained in the range of f. Therefore, the range of f cannot be an interval. □

4.5.2. Monotone inverse theorem. Recall from Section 1.3 that a function has an inverse if and only if the function is injective, that is, one-to-one. Notice that a strictly monotone function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is one-to-one since, for instance, if f is strictly increasing, then $x \neq y$, say $x < y$, implies that $f(x) < f(y)$, which in particular says that $f(x) \neq f(y)$. Thus a strictly monotone function is one-to-one. The last result in this section states that a one-to-one continuous function is automatically strictly monotone. This result makes intuitive sense for if the graph of the function had a dip in it, the function would not pass the so-called "horizontal line test" learned in high school.

THEOREM 4.27 (**Monotone inverse theorem**). *A one-to-one continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is strictly monotone, its range is an interval, and it has a continuous strictly monotone inverse (with the same monotonicity as f).*

PROOF. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a one-to-one continuous function. We shall prove that f is strictly monotone. Fix points $x_0 < y_0$. Then $f(x_0) \neq f(y_0)$ so either $f(x_0) < f(y_0)$ or $f(x_0) > f(y_0)$. For concreteness, assume that $f(x_0) < f(y_0)$; the other case $f(x_0) > f(y_0)$ can be dealt with analogously. We claim that f is strictly increasing. Indeed, if this is not the case, then there exists points $x_1 < y_1$ such that $f(y_1) < f(x_1)$. Now consider the function $g : [0, 1] \to \mathbb{R}$ defined by

$$g(t) = f(ty_0 + (1 - t)y_1) - f(tx_0 + (1 - t)x_1).$$

Since f is continuous, g is continuous, and

$$g(0) = f(y_1) - f(x_1) < 0 \qquad \text{and} \qquad g(1) = f(y_0) - f(x_0) > 0.$$

Hence by the IVP, there is a $c \in [0, 1]$ such that $g(c) = 0$. This implies that $f(a) = f(b)$ where $a = cx_0 + (1 - c)x_1$ and $b = cy_0 + (1 - c)y_1$. Since f is one-to-one, we must have $a = b$; however, this is impossible since $x_0 < y_0$ and $x_1 < y_1$ implies $a < b$. This contradiction shows that f must be strictly monotone.

Now let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous strictly monotone function and let $I = f(\mathbb{R})$. By Theorem 4.26, we know that I is an interval too. We shall prove that $f^{-1} : I \longrightarrow \mathbb{R}$ is also a strictly monotone function; then Theorem 4.26 implies that f^{-1} is continuous. Now suppose, for instance, that f is strictly increasing; we shall prove that f^{-1} is also strictly increasing. If $x < y$ in I, then we can write $x = f(\xi)$ and $y = f(\eta)$ for some ξ and η in I. Since f is increasing, $\xi < \eta$, and hence, $f^{-1}(x) = \xi < \eta = f^{-1}(y)$. Thus, f^{-1} is strictly increasing and our proof is complete. □

Here is a nice application of the monotone inverse theorem.

Example 4.31. Note that $f(x) = x^n$ is monotone on $[0, \infty)$ and strictly increasing. Therefore $f^{-1}(x) = x^{1/n}$ is continuous. In particular, for any $m \in \mathbb{N}$, $g(x) = x^{m/n} = (x^{1/n})^m$ is continuous on $[0, \infty)$ being a composition of the continuous functions f^{-1} and the n-th power. Similarly, $x \mapsto x^{m/n}$ when $m \in \mathbb{Z}$ with $m < 0$ is continuous on $(0, \infty)$. Therefore, for any $r \in \mathbb{Q}$, $x \mapsto x^r$ is continuous on $[0, \infty)$ if $r \geq 0$ and on $(0, \infty)$ if $r < 0$.

EXERCISES 4.5.

1. Prove the following algebraic properties of nondecreasing functions:
   (a) If f and g are nondecreasing, then $f + g$ is nondecreasing.
   (b) If f and g are nondecreasing and nonnegative, then $f\,g$ is nondecreasing.
   (c) Does (b) hold for any (not necessarily nonnegative) nondecreasing functions? Either prove it or give a counterexample.
2. Here is different way to prove that a monotone function has at most countably many discontinuities. Let $f : [a, b] \longrightarrow \mathbb{R}$ be nondecreasing.

(i) Given any finite number $x_1, \ldots, x_k$ of points in (a,b), prove that

$$d(x_1) + \cdots + d(x_k) \leq f(b) - f(a), \qquad \text{where } d(x) := f(x+) - f(x-).$$

(ii) Given any $n \in \mathbb{N}$, prove that there are only a finite number of points $c \in [a,b]$ such that $f(c+) - f(c-) > 1/n$.

(iii) Now prove that f can have at most countably many discontinuities.

3. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a monotone function. Prove that if f happens to also be additive (see Problem 3 in Exercises 4.3), then f is continuous. Thus, any additive monotone function is continuous.

4. In this problem we investigate jump functions. Let $x_1, x_2, \ldots$ be countably many points on the real line and let $c_1, c_2, \ldots$ be nonzero complex numbers such that $\sum c_n$ is absolutely convergent. For $x \in \mathbb{R}$, the functions

$$\varphi_\ell(x) = \sum_{x_n < x} c_n \quad \text{and} \quad \varphi_r(x) = \sum_{x_n \leq x} c_n \tag{4.23}$$

are called a **(left-continuous) jump function** and **(right-continuous) jump function**, respectively. More precisely, $\varphi_\ell(x) := \lim s_n(x)$ and $\varphi_r(x) := \lim t_n(x)$ where

$$s_n(x) := \sum_{k \leq n,\, x_k < x} c_k \quad \text{and} \quad t_n(x) := \sum_{k \leq n,\, x_k \leq x} c_k;$$

thus, e.g. for $s_n(x)$ we only sum over those c_k's such that $k \leq n$ and also $x_k < x$.

(a) Prove that $\varphi_\ell, \varphi_r : \mathbb{R} \longrightarrow \mathbb{C}$ are well-defined for all $x \in \mathbb{R}$ (that is, the two infinite series (4.23) make sense for all $x \in \mathbb{R}$).

(b) If all the c_n's are nonnegative real numbers, prove that φ_ℓ and φ_r are nondecreasing functions on $\mathbb{R}$.

(c) If all the c_n's are nonpositive real numbers, prove that φ_ℓ and φ_r are nonincreasing functions on $\mathbb{R}$.

5. In this problem we prove that φ_r in (4.23) is right-continuous having only jump discontinuities at $x_1, x_2, \ldots$ with the jump at x_n equal to c_n. To this end, let $\varepsilon > 0$. Since $\sum |c_n|$ converges, by Cauchy's criterion for series, we can choose N so that

$$\sum_{n \geq N+1} |c_n| < \varepsilon. \tag{4.24}$$

(i) Prove that for any $\delta > 0$,

$$\varphi_r(x + \delta) - \varphi_r(x) = \sum_{x < x_n \leq x + \delta} c_n.$$

Using (4.24) prove that for $\delta > 0$ sufficiently small, $|\varphi_r(x+\delta) - \varphi_r(x)| < \varepsilon$.

(ii) Prove that for any $\delta > 0$,

$$\varphi_r(x) - \varphi_r(x - \delta) = \sum_{x - \delta < x_n \leq x} c_n.$$

If x is not one of the points $x_1, \ldots, x_N$, using (4.24) prove that for $\delta > 0$ sufficiently small, $|\varphi_r(x) - \varphi_r(x-\delta)| < \varepsilon$.

(iii) If $x = x_k$ for some $1 \leq k \leq N$, prove that $|\varphi_r(x) - \varphi_r(x - \delta) - c_k| < \varepsilon$.

(iv) Finally, prove that φ_r is right-continuous having only jump discontinuities at $x_1, x_2, \ldots$ with the jump at x_n equal to c_n.

6. Prove that φ_ℓ is left-continuous having only jump discontinuities at $x_1, x_2, \ldots$ where the jump at x_n equal to c_n, with the notation given in (4.23).

7. (**Generalized Thomae functions**) In this problem we generalize Thomae's function to arbitrary countable sets. Let $A \subseteq \mathbb{R}$ be a countable set.

(a) Define a nondecreasing function on $\mathbb{R}$ that is discontinuous exactly on A.

(b) Suppose that A is dense. (Dense in defined in Subsection 4.3.3.) Prove that there does not exist a continuous function on $\mathbb{R}$ that is discontinuous exactly on A^c.

4.6. Exponentials, logs, Euler and Mascheroni, and the ζ-function

We now come to a very fun part of real analysis: We apply our work done in the preceding chapters and sections to study the so-called "elementary transcendental functions," the exponential, logarithmic, and trigonometric functions. In particular, we develop the properties of undoubtedly the most important function in all of analysis, the exponential function. We also study logarithms and (complex) powers and derive some of their main properties. For another approach to defining logarithms, see the interesting article [**12**] and for a brief history, [**182**]. In Section 4.7 we define the trigonometric functions.

4.6.1. The exponential function.

Recall that (see Section 3.7) the exponential function is defined by

$$\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad z \in \mathbb{C}.$$

Some properties of the exponential function are found in Theorem 3.31. Here's another important property.

THEOREM 4.28. *The exponential function* $\exp : \mathbb{C} \longrightarrow \mathbb{C}$ *is continuous.*

PROOF. Given any $c \in \mathbb{C}$, using properties *(2)* and *(3)* of Theorem 3.31, we obtain

$$\exp(z) - \exp(c) = \exp(c) \cdot \big[\exp(-c)\exp(z) - 1\big] = \exp(c) \cdot \big[\exp(z-c) - 1\big]. \tag{4.25}$$

Observe that

$$\exp(z-c) - 1 = \sum_{n=0}^{\infty} \frac{(z-c)^n}{n!} - 1 = \sum_{n=1}^{\infty} \frac{(z-c)^n}{n!}.$$

If $|z-c| < 1$, then for $n = 1$, $|z-c|^n = |z-c|$ and for $n > 1$,

$$|z-c|^n = |z-c| \cdot |z-c|^{n-1} < |z-c| \cdot 1 = |z-c|,$$

so by our triangle inequality for series (see Theorem 3.29), we have

$$|z-c| < 1 \quad \Longrightarrow \tag{4.26}$$
$$|\exp(z-c) - 1| \le \sum_{n=1}^{\infty} \frac{|z-c|^n}{n!} < |z-c| \sum_{n=1}^{\infty} \frac{1}{n!} = |z-c| \cdot (e-1).$$

Now let $\varepsilon > 0$. Then choosing $\delta = \min\{1, \varepsilon/(\exp(c)(e-1))\}$, we see that for $|z-c| < \delta$, we have

$$|\exp(z) - \exp(c)| \overset{\text{by (4.25)}}{=} |\exp(c)| \cdot \big|\exp(z-c) - 1\big| \overset{\text{by (4.26)}}{<} |\exp(c)| \cdot |z-c| \cdot (e-1) < \varepsilon.$$

This completes the proof of the theorem. □

An easy induction argument using *(2)* shows that for any complex numbers $z_1, \ldots, z_n$, we have

$$\exp(z_1 + \cdots + z_n) = \exp(z_1) \cdots \exp(z_n).$$

We now restrict the exponential function to real variables $z = x \in \mathbb{R}$:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \qquad x \in \mathbb{R}.$$

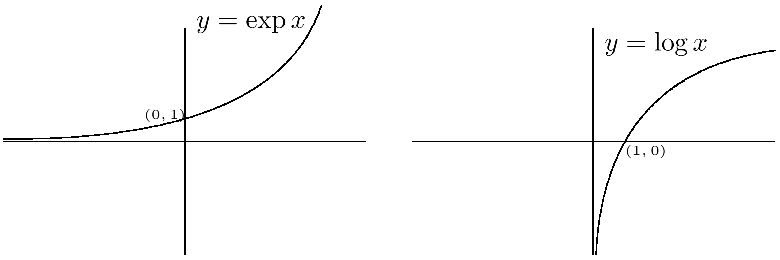


FIGURE 4.10. The graph of $\exp : \mathbb{R} \longrightarrow (0, \infty)$ looks like the graph you learned in high school! Since the exponential function is strictly increasing, it has an inverse function $\exp^{-1}$, which we call the logarithm, $\log : (0, \infty) \longrightarrow \mathbb{R}$.

In particular, the right-hand side, being a sum of real numbers, is a real number, so $\exp : \mathbb{R} \longrightarrow \mathbb{R}$. Of course, this *real* exponential function shares all of the properties *(1) – (4)* as the complex one does. In the following theorem we show that this real-valued exponential function has the increasing/decreasing properties you learned about in elementary calculus; see Figure 4.10.

THEOREM 4.29 (**Properties of the real exponential**). *The real exponential function has the following properties:*

(1) $\exp : \mathbb{R} \longrightarrow (0, \infty)$ *is a strictly increasing continuous bijection. Moreover,* $\lim_{x\to\infty} \exp(x) = \infty$ *and* $\lim_{x\to-\infty} \exp(x) = 0$.

(2) For any $x \in \mathbb{R}$, *we have*

$$1 + x \le \exp(x)$$

with strict inequality for $x \neq 0$, *that is,* $1 + x < \exp(x)$ *for* $x \neq 0$.

PROOF. Observe that

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \ge 1 + x, \qquad x \ge 0,$$

with strict inequalities for $x > 0$. In particular, $\exp(x) > 0$ for $x \ge 0$ and the inequality $\exp(x) \ge 1 + x$ shows that $\lim_{x\to\infty} \exp(x) = \infty$. If $x < 0$, then $-x > 0$, so $\exp(-x) > 0$, and therefore by Property *(3)* of Theorem 3.31,

$$\exp(x) = \frac{1}{\exp(-x)} > 0.$$

Thus, $\exp(x)$ is positive for all $x \in \mathbb{R}$ and recalling Example 4.14, we see that

$$\lim_{x\to-\infty} \exp(x) = \lim_{x\to-\infty} \frac{1}{\exp(-x)} = \lim_{x\to\infty} \frac{1}{\exp(x)} = 0.$$

(As a side note, we can also get $\exp(x) > 0$ for all $x \in \mathbb{R}$ by noting that $\exp(x) = \exp(x/2) \cdot \exp(x/2) = (\exp(x/2))^2$.) If $x < y$, then $y - x > 0$, so $\exp(y - x) \ge 1 + (y - x) > 1$, and thus,

$$\exp(x) < \exp(y - x) \cdot \exp(x) = \exp(y - x + x) = \exp(y).$$

Thus, exp is strictly increasing on $\mathbb{R}$. The continuity property of exp implies that $\exp(\mathbb{R})$ is an interval and then the limit properties of exp imply that this interval must be $(0, \infty)$. Thus, $\exp : \mathbb{R} \longrightarrow (0, \infty)$ is onto (since $\exp(\mathbb{R}) = (0, \infty)$) and injective (since exp is strictly increasing) and therefore is a continuous bijection.

Finally, we verify *(2)*. We already know that $\exp(x) \geq 1 + x$ for $x \geq 0$. If $x \leq -1$, then $1+x \leq 0$ so our inequality is automatically satisfied since $\exp(x) > 0$. If $-1 < x < 0$, then by the series expansion for exp, we have

$$\exp(x) - (1+x) = \left(\frac{x^2}{2!} + \frac{x^3}{3!}\right) + \left(\frac{x^4}{4!} + \frac{x^5}{5!}\right) + \cdots ,$$

where we group the terms in pairs. A typical term in parentheses is of the form

$$\left(\frac{x^{2k}}{(2k)!} + \frac{x^{2k+1}}{(2k+1)!}\right) = \frac{x^{2k}}{(2k)!}\left(1 + \frac{x}{(2k+1)}\right) , \quad k = 1, 2, 3, \ldots .$$

For $-1 < x < 0$, $1+\frac{x}{(2k+1)}$ is positive and so is x^{2k} (being a perfect square). Hence, being a sum of positive numbers, $\exp(x) - (1+x)$ is positive for $-1 < x < 0$. □

The inequality $1 + x \leq \exp(x)$ is quite useful and we will many opportunities to use it in the sequel; see Problem 4 for a nice application to the AGMI.

4.6.2. Existence and properties of logarithms. Since $\exp : \mathbb{R} \longrightarrow (0,\infty)$ is a strictly increasing continuous bijection (so in particular is one-to-one), by the monotone inverse theorem (Theorem 4.27) this function has a strictly increasing continuous bijective inverse $\exp^{-1} : (0,\infty) \longrightarrow \mathbb{R}$. This function is called the **logarithm** function[5] and is denoted by $\log = \exp^{-1}$,

$$\log = \exp^{-1} : (0,\infty) \longrightarrow \mathbb{R}.$$

By definition of the inverse function, log satisfies

$$(4.27) \qquad \exp(\log x) = x, \quad x \in (0,\infty) \qquad \text{and} \qquad \log(\exp x) = x, \quad x \in \mathbb{R}.$$

The logarithm is usually introduced as follows. If $a > 0$, then the unique real number ξ having the property that

$$\exp(\xi) = a$$

is called the **logarithm** of a, where ξ is unique because $\exp : \mathbb{R} \longrightarrow (0,\infty)$ is bijective. Note that $\xi = \log a$ by the second equation in (4.27):

$$\xi = \log(\exp(\xi)) = \log a.$$

THEOREM 4.30. *The logarithm* $\log : (0,\infty) \longrightarrow \mathbb{R}$ *is a strictly increasing continuous bijection. Moreover,* $\lim_{x\to\infty} \log x = \infty$ *and* $\lim_{x\to 0^+} \log x = -\infty$.

PROOF. We already know that log is a strictly increasing continuous bijection. The limit properties of log follow directly from the limit properties of the exponential function in Part *(1)* of the previous theorem, as you can check. □

The following theorem lists some of the well-known properties of log.

THEOREM 4.31 (**Properties of the logarithm**). *The logarithm has the following properties:*

(1) $\exp(\log x) = x$ *and* $\log(\exp x) = x$.
(2) $\log(xy) = \log x + \log y$,

[5] In elementary calculus classes, our logarithm function is denoted by ln and is called the **natural logarithm** function; the notation log usually referring to the "base 10" logarithm. However, in more advanced mathematics, log always refers to the natural logarithm function: *Mathematics is the art of giving the same name to different things. Henri Poincaré (1854–1912). [As opposed to the quotation: Poetry is the art of giving different names to the same thing].*

(3) $\log 1 = 0$. $\log e = 1$.
(4) $\log(x/y) = \log x - \log y$.
(5) $\log x < \log y$ *if and only if* $x < y$.
(6) $\log x > 0$ *if* $x > 1$ *and* $\log x < 0$ *if* $x < 1$.

Proof. We shall leave most of these proofs to the reader. The property *(1)* follows from the fact that exp and log are inverse functions. Consider now the proof of *(2)*. We have

$$\exp(\log(xy)) = xy.$$

On the other hand,

$$\exp(\log x + \log y) = \exp(\log x)\exp(\log y) = xy,$$

so

$$\exp(\log(xy)) = \exp(\log x + \log y).$$

Since exp is one-to-one, we must have $\log(xy) = \log x + \log y$. To prove *(3)*, observe that

$$\exp(0) = 1 = \exp(\log 1),$$

so, because exp is one-to-one, $\log 1 = 0$. Also, since

$$\exp(1) = e = \exp(\log e),$$

by uniqueness, $1 = \log e$. We leave the rest of the properties to the reader. □

4.6.3. Powers and roots of real numbers. Recall that in Section 2.7, we defined the meaning of a^r for $a > 0$ and $r \in \mathbb{Q}$; namely if $r = m/n$ with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then $a^r = \left(\sqrt[n]{a}\right)^m$. We also proved that these rational powers satisfy all the "power rules" that we learned in high school (see Theorem 2.33). We now ask: Can we define a^x for x an arbitrary irrational number. In fact, we shall now define a^z for z an arbitrary *complex* number!

Given any positive real number a and complex number z, we define

$$\boxed{a^z := \exp(z \log a).}$$

The number a is called the **base** and z is called the **exponent**. The astute student might ask: What if $z = k$ is an integer; does this definition of a^k agree with our usual definition of k products of a? What about if $z = p/q \in \mathbb{Q}$, then is the definition of $a^{p/q}$ as $\exp((p/q)\log a)$ in agreement with our previous definition as $\sqrt[q]{a^p}$? We answer these questions and more in the following theorem.

Theorem 4.32 (**Generalized power rules**). *For any real* $a, b > 0$, *we have*

(1) $a^k = a \cdot a \cdots a$ *(k times) for any integer* k.
(2) $e^z = \exp z$ *for all* $z \in \mathbb{C}$.
(3) $\log x^y = y \log x$ *for all* $x, y > 0$.
(4) *For any* $x \in \mathbb{R}, z, w \in \mathbb{C}$,

$$a^z \cdot a^w = a^{z+w}; \quad a^z \cdot b^z = (ab)^z; \quad (a^x)^z = a^{xz}.$$

(5) *If* $z = p/q \in \mathbb{Q}$, *then*

$$a^{p/q} = \sqrt[q]{a^p}.$$

(6) *If* $a > 1$, *then* $x \mapsto a^x$ *is a strictly increasing continuous bijection of* $\mathbb{R}$ *onto* $(0, \infty)$ *and* $\lim_{x\to\infty} a^x = \infty$ *and* $\lim_{x\to-\infty} a^x = 0$. *On the other hand, if* $0 < a < 1$, *then* $x \mapsto a^x$ *is a strictly decreasing continuous bijection of* $\mathbb{R}$ *onto* $(0, \infty)$ *and* $\lim_{x\to\infty} a^x = 0$ *and* $\lim_{x\to-\infty} a^x = \infty$.

(7) If $a, b > 0$ and $x > 0$, then $a < b$ if and only if $a^x < b^x$.

PROOF. By definition of a^k and the additive property of the exponential,

$$a^k = \exp(k\log a) = \exp(\underbrace{\log a + \cdots + \log a}_{k \text{ times}}) = \underbrace{\exp(\log a)\cdots\exp(\log a)}_{k \text{ times}} = \underbrace{a\cdots a}_{k \text{ times}},$$

which proves *(1)*. Since $\log e = 1$, we have

$$e^z = \exp(z\log e) = \exp(z),$$

which is just *(2)*.

To prove *(3)*, observe that

$$\exp(\log(x^y)) = x^y = \exp(y\log x).$$

Since the exponential is one-to-one, we have $\log(x^y) = y\log x$.

If $x \in \mathbb{R}$ and $z, w \in \mathbb{C}$, then the following computations prove *(4)*:

$$\begin{aligned} a^z \cdot a^w = \exp(z\log a)\exp(w\log a) &= \exp(z\log a + w\log a) \\ &= \exp\Big((z+w)\log a\Big) = a^{z+w}; \end{aligned}$$

$$\begin{aligned} a^z \cdot b^z = \exp(z\log a)\exp(z\log b) &= \exp(z\log a + z\log b) \\ &= \exp\Big(z\log(ab)\Big) = (ab)^z, \end{aligned}$$

and

$$(a^x)^z = \exp(z\log a^x) = \exp(xz\log a) = a^{xz}.$$

To prove *(5)*, observe that by the last formula in *(4)*,

$$\Big(a^{p/q}\Big)^p = a^{(p/q)q} = a^p.$$

Therefore, since $a^{p/q} > 0$, by uniqueness of roots (Theorem 2.31), $a^{p/q} = \sqrt[q]{a^p}$.

We leave the reader to verify that since $\exp : \mathbb{R} \longrightarrow \mathbb{R}$ is a strictly increasing bijection with the limits $\lim_{x\to\infty}\exp(x) = \infty$ and $\lim_{x\to-\infty}\exp(x) = 0$, then for any $b > 0$, $\exp(bx)$ is also a strictly increasing continuous bijection of $\mathbb{R}$ onto $(0,\infty)$ and $\lim_{x\to\infty}\exp(bx) = \infty$ and $\lim_{x\to-\infty}\exp(bx) = 0$. On the other hand, if $b < 0$, say $b = -c$ where $c > 0$, then these properties are reversed: $\exp(-cx)$ is a strictly decreasing continuous bijection of $\mathbb{R}$ onto $(0,\infty)$ and $\lim_{x\to\infty}\exp(-cx) = 0$ and $\lim_{x\to-\infty}\exp(-cx) = \infty$. With this discussion in mind, note that if $a > 1$, then $\log a > 0$ (Property *(6)* of Theorem 4.31), so $a^x = \exp(x\log a) = \exp(bx)$ has the required properties in *(6)*; if $0 < a < 1$, then $\log a < 0$, so $a^x = \exp(x\log a) = \exp(-cx)$, where $c = -\log a > 0$, has the required properties in *(6)*.

Finally, to verify *(7)*, observe that for $a, b > 0$ and $x > 0$, using the fact that log and exp are strictly increasing, we obtain

$$\begin{aligned} a < b \quad &\Longleftrightarrow \quad \log a < \log b \quad \Longleftrightarrow \quad x\log a < x\log b \\ &\Longleftrightarrow \quad a^x = \exp(x\log a) < \exp(x\log b) = b^x. \end{aligned}$$

□

Example 4.32. Using Tannery's theorem, we shall prove the pretty formula

$$\boxed{\frac{e}{e-1} = \lim_{n\to\infty}\left\{\left(\frac{n}{n}\right)^n + \left(\frac{n-1}{n}\right)^n + \left(\frac{n-2}{n}\right)^n + \cdots + \left(\frac{1}{n}\right)^n\right\}.}$$

To prove this, we write the right-hand side as

$$\lim_{n\to\infty}\left\{\Big(\frac{n}{n}\Big)^n + \Big(\frac{n-1}{n}\Big)^n + \cdots + \Big(\frac{1}{n}\Big)^n\right\} = \lim_{n\to\infty}\sum_{k=0}^{\infty} a_k(n),$$

where $a_k(n) := 0$ for $k \geq n$ and for $0 \leq k \leq n-1$,

$$a_k(n) := \Big(\frac{n-k}{n}\Big)^n = \Big(1 - \frac{k}{n}\Big)^n.$$

Observe that

$$\lim_{n\to\infty} a_k(n) = \lim_{n\to\infty}\Big(1-\frac{k}{n}\Big)^n = e^{-k}$$

exists. Also, for $k \leq n-1$,

$$|a_k(n)| = \Big(1-\frac{k}{n}\Big)^n \leq \Big(e^{-k/n}\Big)^n = e^{-k},$$

where we used that $1 + x \leq e^x$ for all $x \in \mathbb{R}$ from Theorem 4.29. Since $a_k(n) = 0$ for $k \geq n$, it follows that $|a_k(n)| \leq M_k$ for all k, n where $M_k = e^{-k}$. Since $e^{-1} < 1$, by the geometric series test, $\sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty}(e^{-1})^k < \infty$. Hence by Tannery's theorem, we have

$$\lim_{n\to\infty}\left\{\Big(\frac{n}{n}\Big)^n + \Big(\frac{n-1}{n}\Big)^n + \cdots + \Big(\frac{1}{n}\Big)^n\right\}$$
$$= \lim_{n\to\infty}\sum_{k=0}^{\infty} a_k(n) = \sum_{k=0}^{\infty}\lim_{n\to\infty} a_k(n) = \sum_{k=0}^{\infty} e^{-k} = \frac{1}{1-1/e} = \frac{e}{e-1}.$$

Example 4.33. Here's a **Puzzle:** Do there exist *rational* numbers α and β such that α^β is *irrational*? You should be able to answer this in the affirmative! Here's a harder question [**108**]: Do there exist *irrational* numbers α and β such that α^β is *rational*? Here's a very cool argument to the affirmative. Consider $\alpha = \sqrt{2}$ and $\beta = \sqrt{2}$, both of which are irrational. Then there are two cases: either α^β rational or irrational. If α^β is rational, then we have answered our question in the affirmative. However, in the case that $\alpha' := \alpha^\beta$ is irrational, then by our rule *(4)* of exponents,

$$(\alpha')^\beta = \Big(\alpha^\beta\Big)^\beta = \alpha^{\beta^2} = \sqrt{2}^2 = 2$$

is rational, so we have answered our question in the affirmative in this case as well. Do there exist *irrational* numbers α and β such that α^β is *irrational*? For the answer, see Problem 6.

4.6.4. The Riemann zeta function. The last two subsections are applications of what we've learned about exponentials, logs, and powers. We begin with the Riemann zeta-function, which is involved in one of the most renowned unsolved problems in all of mathematics: *The Riemann hypothesis.* If you want to be famous and earn one million dollars too, just prove the Riemann hypothesis (see http://www.claymath.org/millennium/ and [**53**] for hints on how one may try to solve this conjecture); for now, our goal is simply to introduce this function. Actually, this function is simply a "generalized p-series" where instead of using p, a rational number, we use a complex number:

$$\boxed{\zeta(z) := \sum_{n=1}^{\infty}\frac{1}{n^z} = 1 + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \cdots .}$$

THEOREM 4.33 (**The Riemann zeta function**). *The Riemann zeta-function converges absolutely for all $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$.*

PROOF. Let p be an arbitrary rational number with $p > 1$; then we just have to prove that $\zeta(z)$ converges absolutely for all $z \in \mathbb{C}$ with $\operatorname{Re} z \geq p$. To see this, let $z = x + iy$ with $x \geq p$ and observe that $n^z = e^{z \log n} = e^{x \log n} \cdot e^{iy \log n}$. In Problem 1d you'll prove that $|e^{i\theta}| = 1$ for any real θ, so $|e^{iy \log n}| = 1$ and hence,

$$|n^z| = |e^{x \log n} \cdot e^{iy \log n}| = e^{x \log n} \geq e^{p \log n} = n^p.$$

Therefore, $\big|1/n^z\big| \leq 1/n^p$, so by comparison with the p-series $\sum 1/n^p$, it follows that $\sum |1/n^z|$ converges. This completes our proof. □

The ζ-function has profound implications to prime numbers; see Section 7.6.

4.6.5. The Euler-Mascheroni constant.

The constant

$$\boxed{\gamma := \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right)}$$

is called the **Euler-Mascheroni constant**. This constant was calculated to 16 digits by Euler in 1781, who used the notation C for γ. The symbol γ was first used by Lorenzo Mascheroni (1750–1800) in 1790 when he computed γ to 32 decimal places, although only the first 19 places were correct (cf. [**96**, pp. 90–91]). To prove that the limit on the right of γ exists, consider the sequence

$$\gamma_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n, \qquad n = 2, 3, \ldots.$$

We shall prove that γ_n is nonincreasing and bounded below and hence the Euler-Mascheroni constant is defined. In our proof, we shall see that γ is between 0 and 1; the exact value in base 10 is $\gamma = .5772156649\ldots$. Here's a mnemonic to remember the digits of γ [**236**]:

(4.28) *These numbers proceed to a limit Euler's subtle mind discerned.*

The number of letters in each word represents a digit of γ; e.g. "These" represents 5, "numbers" 7, etc. The sentence (4.28) gives ten digits of γ: .5772156649. By the way, it is not known[6] whether γ is rational or irrational, let alone transcendental!

To prove that $\{\gamma_n\}$ is a bounded monotone sequence, we shall need the following inequality proved in Section 3.3 (see (3.28)):

$$\left(\frac{n+1}{n}\right)^n < e < \left(\frac{n+1}{n}\right)^{n+1} \qquad \text{for all } n \in \mathbb{N}.$$

Taking the logarithm of both sides of the first inequality and using the fact that log is strictly increasing implies, we get

$$n\big(\log(n+1) - \log n\big) = n \log\left(\frac{n+1}{n}\right) < \log e = 1,$$

and doing the same thing to the second inequality gives

$$1 = \log e < (n+1) \log\left(\frac{n+1}{n}\right) = (n+1)\big(\log(n+1) - \log n\big).$$

[6] *Unfortunately what is little recognized is that the most worthwhile scientific books are those in which the author clearly indicates what he does not know; for an author most hurts his readers by concealing difficulties. Evariste Galois (1811–1832).* [**188**].

Combining these two inequalities, we obtain

$$\frac{1}{n+1} < \log(n+1) - \log n < \frac{1}{n}. \tag{4.29}$$

Using the definition of γ_n and first inequality in (4.29), we see that

$$\gamma_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n = \gamma_{n+1} - \frac{1}{n+1} + \log(n+1) - \log n > \gamma_{n+1},$$

so the sequence $\{\gamma_n\}$ is strictly decreasing. In particular, $\gamma_n < \gamma_1 = 1$ for all n. We now show that γ_n is bounded below by zero. We already know that $\gamma_1 = 1 > 0$. Using the second inequality in (4.29) with $n = 2, n = 3, \ldots, n = n$, we obtain

$$\begin{aligned}\gamma_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n &> 1 + \Big(\log 3 - \log 2\Big) + \Big(\log 4 - \log 3\Big) \\ + \Big(\log 5 - \log 4\Big) + \cdots + \Big(\log n - \log(n-1)\Big) &+ \Big(\log(n+1) - \log n\Big) - \log n \\ &= 1 - \log 2 + \log(n+1) - \log n > 1 - \log 2 > 0.\end{aligned}$$

Here, we used that $\log 2 < 1$ because $2 < e$. Thus, $\{\gamma_n\}$ is strictly decreasing and bounded below by $1 - \log 2 > 0$, so γ is is well-defined and $0 < \gamma < 1$.

We can now show that the value of the alternating harmonic series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + - \cdots$$

is $\log 2$. Indeed, since

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right),$$

we see that

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{2n} - \log 2n\right)$$

and

$$\gamma = \lim_{n\to\infty} 2\left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\right) - \log n.$$

Subtracting, we obtain

$$0 = \lim_{n\to\infty} \left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots - \frac{1}{2n}\right) - \log 2,$$

which proves that

$$\boxed{\log 2 = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + - \cdots .}$$

Using a similar technique, one can find series representations for $\log 3$; see Problem 7. Using the above formula for $\log 2$, in Problem 7 you are asked to derive the following striking expression:

$$\boxed{2 = \frac{e^1}{e^{1/2}} \cdot \frac{e^{1/3}}{e^{1/4}} \cdot \frac{e^{1/5}}{e^{1/6}} \cdot \frac{e^{1/7}}{e^{1/8}} \cdot \frac{e^{1/9}}{e^{1/10}} \cdots .} \tag{4.30}$$

Exercises 4.6.

1. Establish the following properties of exponential functions.
   (a) If $z_n \to z$ and $a_n \to a$ (with z_n, z complex and $a_n, a > 0$), then $a_n^{z_n} \to a^z$.

(b) If $a, b > 0$, then for any $x < 0$, $a < b$ if and only if $a^x > b^x$.
(c) If $a, b > 0$, then for any complex number z, $a^{-z} = 1/a^z$ and $(a/b)^z = a^z/b^z$.
(d) Prove that for any $x \in \mathbb{R}$, $|e^{ix}| = 1$.

2. Let $a \in \mathbb{R}$ with $a \neq 0$ and define $f(x) = x^a$.
(a) If $a > 0$, prove that $f : [0, \infty) \longrightarrow \mathbb{R}$ is continuous, strictly increasing, $\lim_{x \to 0+} f = 0$, and $\lim_{x \to \infty} f = \infty$.
(b) If $a < 0$, prove that $f : (0, \infty) \longrightarrow \mathbb{R}$ is continuous, strictly decreasing, $\lim_{x \to 0+} f = \infty$, and $\lim_{x \to \infty} f = 0$.

3. Establish the following limit properties of the exponential function.
(a) Show that for any natural number n and for any $x \in \mathbb{R}$ with $x > 0$ we have
$$e^x > \frac{x^{n+1}}{(n+1)!}.$$
Use this inequality to prove that for any natural number n,
$$\lim_{x \to \infty} \frac{x^n}{e^x} = 0.$$
(b) Using (a), prove that for any $a \in \mathbb{R}$ with $a > 0$, however large, we have
$$\lim_{x \to \infty} \frac{x^a}{e^x} = 0.$$
It follows that e^x grows faster than any power (no matter how large) of x. This limit is usually derived in elementary calculus using L'Hospital's rule.

4. Let $a_1, \ldots, a_n$ be nonnegative real numbers. Recall from Problem 7 in Exercises 2.2 that the **arithmetic-geometric mean inequality** (AGMI) is the iequality
$$(a_1 \cdot a_2 \cdots a_n)^{1/n} \leq \frac{a_1 + \cdots + a_n}{n}.$$
Prove this inequality by setting $a = (a_1 + \cdots + a_n)/n$, $x_k = -1 + a_k/a$ (so that $a_k/a = 1 + x_k$) for $k = 1, \ldots, n$, and using the inequality $1 + x \leq e^x$.

5. For any $x > 0$, derive the following remarkable formula:
$$\log x = \lim_{n \to \infty} n\big(\sqrt[n]{x} - 1\big) \qquad (\textbf{Halley's formula}),$$
named after the famous Edmond Halley (1656–1742) of Halley's comet. Suggestion: Write $\sqrt[n]{x} = e^{\log x/n}$ and write $e^{\log x/n}$ as a series in $\log x/n$.

6. (Cf. [**108**]) **Puzzle:** Do there exist *irrational* numbers α and β such that α^β is *irrational*? Suggestion: Consider α^β and $\alpha^{\beta'}$ where $\alpha = \beta = \sqrt{2}$ and $\beta' = \sqrt{2} + 1$.

7. In this fun problem, we derive some interesting formulas.
(a) Prove that
$$\gamma = \sum_{n=1}^{\infty} \left[\frac{1}{n} - \log\left(1 + \frac{1}{n}\right)\right] = 1 + \sum_{n=2}^{\infty} \left[\frac{1}{n} + \log\left(1 - \frac{1}{n}\right)\right]$$
$$= 1 + \sum_{n=1}^{\infty} \left[\frac{1}{n+1} + \log\left(1 + \frac{1}{n}\right)\right],$$
where γ is the Euler-Mascheroni constant. Suggestion: Think telescoping series.
(b) Using a similar technique on how we derived our formula for $\log 2$, prove that
$$\log 3 = 1 + \frac{1}{2} - \frac{2}{3} + \frac{1}{4} + \frac{1}{5} - \frac{2}{6} + \frac{1}{7} + \frac{1}{8} - \frac{2}{9} + + - \cdots$$
Can you find a series representation for $\log 4$?
(c) Define $a_n = \frac{e^1}{e^{1/2}} \cdots \frac{e^{2n-1}}{e^{2n}}$. Prove that $2 = \lim a_n$.

8. Following Greenstein [**86**] (cf. [**42**]) we establish a "well-known" limit from calculus, but without using calculus!
(i) Show that $\log x < x$ for all $x > 0$.
(ii) Show that $(\log x)/x < 2/x^{1/2}$ for $x > 0$. Suggestion: $\log x = 2 \log x^{1/2}$.

(iii) Show that

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

This limit is usually derived in elementary calculus using L'Hospital's rule.

(iv) Now let $a\in\mathbb{R}$ with $a>0$. Generalizing the above argument, prove that

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

Thus, $\log x$ grows slower than any power (no matter how small) of x.

9. In this problem we get an inequality for $\log(1+x)$ and use it to obtain a nice formula.
   (i) Prove that for all $x\in[0,1]$, we have $e^{\frac{1}{2}x}\leq 1+x$. Conclude that for all $x\in[0,1]$, we have $\log(1+x)\geq x/2$.
   (ii) Using Tannery's theorem, prove that

$$\zeta(2)=\lim_{n\to\infty}\left\{\frac{1}{n^2\log\left(1+\frac{1}{n^2}\right)}+\frac{1}{n^2\log\left(1+\frac{2^2}{n^2}\right)}+\frac{1}{n^2\log\left(1+\frac{3^2}{n^2}\right)}+\cdots+\frac{1}{n^2\log\left(1+\frac{n^2}{n^2}\right)}\right\}.$$

10. In high school you probably learned logarithms with other "bases" besides e. Let $a\in\mathbb{R}$ with $a>0$ and $a\neq 1$. For any $x>0$, we define

$$\log_a x:=\frac{\log x}{\log a},$$

called the **logarithm of** x **to the base** a. Note that if $a=e$, then $\log_e=\log$, our usual logarithm. Here are some of the well-known properties of $\log_a$.

(a) Prove that $x\mapsto\log_a x$ is the inverse function of $x\mapsto a^x$.

(b) Prove that for any $x,y>0$, $\log_a xy=\log_a x+\log_a y$.

(c) Prove that if $b>0$ with $b\neq 1$ is another base, then for any $x>0$,

$$\log_a x=\left(\frac{\log b}{\log a}\right)\log_b x\qquad(\textbf{Change of base formula}).$$

11. Part (a) of this problem states that a "function which looks like an exponential function is an exponential function," while (b) says the same for the logarithm function.

(a) Let $f:\mathbb{R}\longrightarrow\mathbb{R}$ satisfy $f(x+y)=f(x)\,f(y)$ for all $x,y\in\mathbb{R}$; see Problem 4 in Exercises 4.3. Assume that f is not the zero function. Prove that if f is continuous, then

$$f(x)=a^x\quad\text{for all } x\in\mathbb{R}\text{, where } a=f(1).$$

Suggestion: Show that $f(x)>0$ for all x. Now there are a couple ways to proceed. One way is to first prove that $f(r)=(f(1))^r$ for all rational r (to prove this you do not require the continuity assumption). This second way is to define $h(x)=\log f(x)$. Prove that h is linear and then apply Problem 3 in Exercises 4.3.

(b) Let $g:(0,\infty)\longrightarrow\mathbb{R}$ satisfy $g(x\cdot y)=g(x)+g(y)$ for all $x,y>0$. Prove that if g is continuous, then there exists a unique real number c such that

$$g(x)=c\log x\quad\text{for all } x\in(0,\infty).$$

12. (**Exponentials the "old fashion way"**) Fix $a>0$ and $x\in\mathbb{R}$. In this section we defined $a^x:=\exp(x\log a)$ However, in this problem we shall define real powers the "old fashion way" via rational sequences. Henceforth we only assume knowledge of rational powers and we proceed to define them for real powers.
   (i) Let $\{r_n\}$ be a sequence of rational numbers converging to zero. From Section 3.1 we know that $a^{1/n}\to 1$ and $a^{-1/n}=(a^{-1})^{1/n}\to 1$. Let $\varepsilon>0$ and fix $m\in\mathbb{N}$ such that $1-\varepsilon<a^{\pm 1/m}<1+\varepsilon$. Show that if $|r_n|<1/m$, then $1-\varepsilon<a^{r_n}<1+\varepsilon$. Conclude that $a^{r_n}\to 1$. (See Problem 3 in Exercises 3.1 for another proof.) Suggestion: Recall that any rational $p<q$ and real $b>1$, we have $b^p<b^q$.

(ii) Let $\{r_n\}$ be a sequence of rational numbers converging to x. Prove that $\{a^{r_n}\}$ is a Cauchy sequence, hence it converges to a real number, say ξ. We define $a^x = \xi$. Prove that this definition makes sense; that is, if $\{r_n'\}$ is any other sequence of rational numbers converging to x, then $\{a^{r_n'}\}$ also converges to ξ.

(iii) Prove that if $x = n \in \mathbb{N}$, then $a^x = a \cdot a \cdots a$ where there are n a's multiplied together. Also prove that $a^{-x} = 1/a^n$ and if $x = n/m \in \mathbb{Q}$, then $a^x = \sqrt[m]{a^n}$. Thus, our new definition of powers agrees with the old definition. Finally, show that for $x, y \in \mathbb{R}$,

$$a^x \cdot a^y = a^{x+y}; \quad a^x \cdot b^x = (ab)^x; \quad (a^x)^y = a^{xy}.$$

13. (**Logarithms the "old fashion way"**) In this problem we define the logarithm the "old fashion way" using rational sequences. In this problem we assume knowledge of real powers as presented in the previous problem. Fix $a > 0$.

(i) Prove that it is possible to define unique integers $a_0, a_1, a_2, \ldots$ inductively with $0 \le a_k \le 9$ for $k \ge 1$ such that if x_n and y_n are the rational numbers

$$x_n = a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_{n-1}}{10^{n-1}} + \frac{a_n}{10^n}$$

and

$$y_n = a_0 + \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_{n-1}}{10^{n-1}} + \frac{a_n + 1}{10^n},$$

then

$$e^{x_n} \le a < e^{y_n}. \tag{4.31}$$

Suggestion: Since $e > 1$, we know that for $r \in \mathbb{Q}$, we have the limits $e^r \to \infty$, respectively 0, as $r \to \infty$, respectively $r \to -\infty$.

(ii) Prove that both sequences $\{x_n\}$ and $\{y_n\}$ converge to the same value, call it L. Show that $e^L = a$ where e^L is defined by means of the previous problem. Of course, L is just the logarithm of a defined in this section.

14. (**The Euler-Mascheroni constant II**) In this problem we prove that the Euler-Mascheroni constant constant exists following [**44**]. Consider the sequence

$$a_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n-1} - \log n, \qquad n = 2, 3, \ldots.$$

We shall prove that a_n is nondecreasing and bounded and hence $\lim a_n$ exists.

(i) Assuming that the limit $\lim a_n$ exists, prove that the limit defining the Euler-Mascheroni constant also exists and equals $\lim a_n$.

(ii) Using the inequalities in (3.28), prove that

$$1 < \frac{e^{1/n}}{(n+1)/n} \qquad \text{and} \qquad \frac{e^{1/n}}{(n+1)/n} < e^{\frac{1}{n(n+1)}}. \tag{4.32}$$

(iii) Prove that for each $n \ge 2$,

$$a_n = \log\left(\frac{e^1}{2/1} \cdot \frac{e^{1/2}}{3/2} \cdots \frac{e^{1/(n-1)}}{n/(n-1)}\right).$$

(iv) Using (c) and the inequalities in (4.32), prove that $\{a_n\}$ is strictly increasing such that $0 < a_n < 1$ for all n. Conclude that $\lim a_n$ exists.

15. (**The Euler-Mascheroni constant III**) We prove that Euler-Mascheroni constant exists following [**16**]. For each $k \in \mathbb{N}$, define $a_k := e\left(1 + \frac{1}{k}\right)^{-k}$ so that $e = a_k \left(1 + \frac{1}{k}\right)^k$.

(i) Prove that

$$\frac{1}{k} = \log(a_k^{1/k}) + \log(k+1) - \log k.$$

(ii) Prove that

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log(n+1) = \log\left(a_1\, a_2^{1/2}\, a_3^{1/3} \cdots a_n^{1/n}\right).$$

(iii) Prove that the sequence $\left\{\log\left(a_1\, a_2^{1/2} \cdots a_n^{1/n}\right)\right\}$ is nondecreasing. Conclude that if this sequence is bounded, then Euler's constant exists.

(iv) Prove that

$$\begin{aligned}\log\left(a_1\, a_2^{1/2} \cdots a_n^{1/n}\right) &= \log a_1 + \frac{1}{2}\log a_2 + \cdots + \frac{1}{n}\log a_n \\ &< \log\left(1+\frac{1}{1}\right) + \frac{1}{2}\log\left(1+\frac{1}{2}\right) + \cdots + \frac{1}{n}\log\left(1+\frac{1}{n}\right) \\ &< \frac{1}{1} + \frac{1}{2}\cdot\frac{1}{2} + \cdots + \frac{1}{n}\cdot\frac{1}{n}.\end{aligned}$$

(v) Since of the reciprocals of the squares converges, conclude that the sequence $\left\{\log\left(a_1\, a_2^{1/2} \cdots a_n^{1/n}\right)\right\}$ is bounded.

4.7. The trig functions, the number π, and which is larger, π^e or e^π?

In high school we learned about sine and cosine using geometric intuition based on either triangles or the unit circle. (For this point of view, see the interesting paper [**211**].) In this section we introduce these function from a purely analytic framework and we prove that these functions have all the properties you learned in high school. In high school we also learned about the number π,[7] again using geometric intuition. In this section we *define* π rigourously using analysis without any geometry. However, we do prove that π has all the geometric properties you think it does.

4.7.1. The trigonometric and hyperbolic functions. We define cosine and sine as the functions $\cos : \mathbb{C} \longrightarrow \mathbb{C}$ and $\sin : \mathbb{C} \longrightarrow \mathbb{C}$ defined by the equations

$$\boxed{\cos z := \frac{e^{iz}+e^{-iz}}{2}, \qquad \sin z := \frac{e^{iz}-e^{-iz}}{2i}.}$$

In particular, both of these functions are continuous functions, being constant multiples of a sum and difference, respectively, of the continuous functions $e^{iz} = \exp(iz)$ and $e^{-iz} = \exp(-iz)$. From these formulas, we see that $\cos 0 = 1$ and $\sin 0 = 0$; other "well-known" values of sine and cosine are discussed in the problems. Multiplying the equation for $\sin z$ by i and then adding this equation to $\cos z$, the e^{-iz} terms cancel and we get $\cos z + i \sin z = e^{iz}$. This equation is the famous **Euler's identity**:

$$\boxed{e^{iz} = \cos z + i \sin z. \qquad (\textbf{Euler's identity})}$$

This formula provides a very easy proof of **de Moivre's formula**, named after its discoverer Abraham de Moivre (1667–1754),

$$\boxed{(\cos z + i \sin z)^n = \cos nz + i \sin nz, \quad z \in \mathbb{C}, \qquad (\textbf{de Moivre's formula}),}$$

which is given much attention in elementary mathematics and is usually only stated when $z = \theta$, a real variable. Here is the one-line proof:

$$(\cos z + i\sin z)^n = \left(e^{iz}\right)^n = \underbrace{e^{iz}\cdot e^{iz}\cdots e^{iz}}_{n \text{ terms}} = e^{inz} = \cos nz + i \sin nz.$$

[7] *"Cosine, secant, tangent, sine, 3.14159; integral, radical, u dv, slipstick, sliderule, MIT!" MIT cheer.*

In the following theorem, we adopt the standard notation of writing $\sin^2 z$ for $(\sin z)^2$, etc.[8] Here are some well-known trigonometric identities that you memorized in high school, now proved from the basic definitions and even for complex variables.

THEOREM 4.34 (**Basic properties of cosine and sine**). *Cosine and sine are continuous functions on $\mathbb{C}$. In particular, restricting to real values, they define continuous functions on $\mathbb{R}$. Moreover, for any complex numbers z and w,*

(1) $\cos(-z) = \cos z$, $\sin(-z) = -\sin z$,

(2) $\cos^2 z + \sin^2 z = 1$, ***(Pythagorean identity)***

(3) Addition formulas:

$$\cos(z+w) = \cos z \cos w - \sin z \sin w, \quad \sin(z+w) = \sin z \cos w + \cos z \sin w,$$

(4) Double angle formulas:

$$\cos(2z) = \cos^2 z - \sin^2 z = 2\cos^2 z - 1 = 1 - 2\sin^2 z,$$
$$\sin(2z) = 2\cos z \sin z.$$

(5) Trigonometric series:[9]

$$\boxed{\cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}, \qquad \sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!},} \tag{4.33}$$

where the series converge absolutely.

PROOF. We shall leave some of this proof to the reader. Note that *(1)* follows directly from the definition of cosine and sine. Consider the addition formula:

$$\begin{aligned}\cos z \cos w - \sin z \sin w &= \left(\frac{e^{iz}+e^{-iz}}{2}\right)\left(\frac{e^{iw}+e^{-iw}}{2}\right) \\ &\qquad - \left(\frac{e^{iz}-e^{-iz}}{2i}\right)\left(\frac{e^{iw}-e^{-iw}}{2i}\right) \\ &= \frac{1}{4}\Big\{e^{i(z+w)} + e^{i(z-w)} + e^{-i(z-w)} + e^{-i(z+w)} \\ &\qquad + e^{i(z+w)} - e^{i(z-w)} - e^{-i(z-w)} + e^{-i(z+w)}\Big\} \\ &= \frac{e^{i(z+w)} + e^{-i(z+w)}}{2} = \cos(z+w).\end{aligned}$$

Taking $w = -z$ and using *(1)* we get the Pythagorean identity:

$$1 = \cos 0 = \cos(z - z) = \cos z \cos(-z) - \sin z \sin(-z) = \cos^2 z + \sin^2 z.$$

[8] Sin$^2\phi$ *is odious to me, even though Laplace made use of it; should it be feared that* $\sin^2\phi$ *might become ambiguous, which would perhaps never occur, or at most very rarely when speaking of* $\sin(\phi^2)$*, well then, let us write* $(\sin\phi)^2$*, but not* $\sin^2\phi$*, which by analogy should signify* $\sin(\sin\phi)$*. Carl Friedrich Gauss (1777–1855).*

[9] In elementary calculus, these series are usually derived via Taylor series and are usually attributed to Sir Isaac Newton (1643–1727) who derived them in his paper "De Methodis Serierum et Fluxionum" (Method of series and fluxions) written in 1671. However, it is interesting to know that these series were first discovered *hundreds* of years earlier by Madhava of Sangamagramma (1350–1425), a mathematicians from the Kerala state in southern India!

We leave the double angle formulas to the reader. To prove *(5)*, we use the power series for the exponential to compute

$$e^{iz} + e^{-iz} = \sum_{n=0}^{\infty} \frac{i^n\, z^n}{n!} + \sum_{n=0}^{\infty} \frac{(-1)^n\, i^n\, z^n}{n!}.$$

The terms when n is odd cancel, so

$$2\cos z = e^{iz} + e^{-iz} = 2\sum_{n=0}^{\infty} \frac{i^{2n}\, z^{2n}}{(2n)!} = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!},$$

where we used the fact that $i^{2n} = (i^2)^n = (-1)^n$. This series converges absolutely since it is the sum of two absolutely convergent series. The series expansion for $\sin z$ is proved in a similar manner. □

From the series expansion for sin it is straightforward to *prove* the following limit from elementary calculus (but now for complex numbers):

$$\boxed{\lim_{z\to 0} \frac{\sin z}{z} = 1;}$$

see Problem 3. Of course, from the identities in Theorem 4.34, one can derive other identities such as the so-called *half-angle formulas*:

$$\cos^2 z = \frac{1 + \cos 2z}{2}, \quad \sin^2 z = \frac{1 - \cos 2z}{2}.$$

The other trigonometric functions are defined in terms of sin and cos in the usual manner:

$$\boxed{\begin{array}{ll} \tan z = \dfrac{\sin z}{\cos z}, & \cot z = \dfrac{1}{\tan z} = \dfrac{\cos z}{\sin z} \\[2ex] \sec z = \dfrac{1}{\cos z}, & \csc z = \dfrac{1}{\sin z}, \end{array}}$$

and are called the **tangent**, **cotangent**, **secant**, and **cosecant**, respectively. Note that these functions are only defined for those complex z for which the expressions make sense, e.g. $\tan z$ is defined only for those z such that $\cos z \neq 0$. The extra trig functions satisfy the same identities that you learned in high school, for example, for any complex numbers z, w, we have

$$\tan(z + w) = \frac{\tan z + \tan w}{1 - \tan z\, \tan w}, \tag{4.34}$$

for those z, w such that the denominator is not zero. Setting $z = w$, we see that

$$\tan 2z = \frac{2\tan z}{1 - \tan^2 z}.$$

In Problem 4 we ask you to prove (4.34) and other identities.

Before baking our π, we quickly define the hyperbolic functions. For any complex number z, we define

$$\boxed{\cosh z := \frac{e^z + e^{-z}}{2}, \qquad \sinh z := \frac{e^z - e^{-z}}{2};}$$

these are called the **hyperbolic cosine** and **hyperbolic sine**, respectively. There are hyperbolic tangents, secants, etc ... defined in the obvious manner. Observe that, by definition, $\cosh z = \cos iz$ and $\sinh z = -i \sin iz$, so after substituting iz for z in the series for cos and sin, we obtain

$$\boxed{\cosh z = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!}, \qquad \sinh z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}.}$$

These functions are intimately related to the trig functions and share many of the same properties; see Problem 8.

4.7.2. The number π and some trig identities. Setting $z = x \in \mathbb{R}$ into the series (4.33), we obtain the formulas learned in elementary calculus:

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \qquad \sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

In particular, $\cos, \sin : \mathbb{R} \longrightarrow \mathbb{R}$. In the following lemma and theorem we shall consider these real-valued functions instead of the more general complex versions. The following lemma is the key result needed to define π.

Lemma 4.35. *Sine and cosine have the following properties on* $[0, 2]$*:*

(1) sin *is nonnegative on* $[0, 2]$ *and positive on* $(0, 2]$*;*
(2) $\cos : [0, 2] \longrightarrow \mathbb{R}$ *is strictly decreasing with* $\cos 0 = 1$ *and* $\cos 2 < 0$.

Proof. Since

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x\left(1 - \frac{x^2}{2 \cdot 3}\right) + \frac{x^5}{5!}\left(1 - \frac{x^2}{6 \cdot 7}\right) + \cdots$$

and each term in the series is positive for $0 < x < 2$, we have $\sin x > 0$ for all $0 < x < 2$ and $\sin 0 = 0$.

Since

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots,$$

we have

$$\cos 2 = 1 - \frac{2^2}{2!} + \frac{2^4}{4!} - \left(\frac{2^6}{6!} - \frac{2^8}{8!}\right) - \left(\frac{2^{10}}{10!} - \frac{2^{12}}{12!}\right) - \cdots.$$

All the terms in parentheses are positive because for $k \geq 2$, we have

$$\frac{2^k}{k!} - \frac{2^{k+2}}{(k+2)!} = \frac{2^k}{k!}\left(1 - \frac{4}{(k+1)(k+2)}\right) > 0.$$

Therefore,

$$\cos 2 < 1 - \frac{2^2}{2!} + \frac{2^4}{4!} = -\frac{1}{3} < 0.$$

We now show that cos is strictly decreasing on $[0, 2]$. Since cos is continuous, by Theorem 4.27 if we show that cos is one-to-one on $[0, 2]$, then we can conclude that cos is strictly monotone on $[0, 2]$; then $\cos 0 = 1$ and $\cos 2 < 0$ tells us that cos must be strictly decreasing. Suppose that $0 \leq x \leq y \leq 2$ and $\cos x = \cos y$; we shall prove that $x = y$. We already know that sin is nonnegative on $[0, 2]$, so the identity

$$\sin^2 x = 1 - \cos^2 x = 1 - \cos^2 y = \sin^2 y$$

implies that $\sin x = \sin y$. Therefore,

$$\sin(y - x) = \sin y \cos x - \cos y \sin x = \sin x \cos x - \cos x \sin x = 0,$$

and using that $0 \le y - x \le 2$ and *(1)*, we get $y - x = 0$. Hence, $x = y$, so cos is one-to-one on $[0, 2]$, and our proof is complete. □

We now define the real number π.

THEOREM 4.36 (**Definition of π**). *There exists a unique real number, denoted by the Greek letter π, having the following two properties:*

(1) $3 < \pi < 4$,

(2) $\cos(\pi/2) = 0$.

Moreover, $\cos x > 0$ *for* $0 < x < \pi/2$.

PROOF. By our lemma, we know that $\cos : [0, 2] \longrightarrow \mathbb{R}$ is strictly decreasing with $\cos 0 = 1$ and $\cos 2 < 0$, so by the intermediate value theorem and the fact that cos is strictly decreasing, there is a unique point $0 < c < 2$ such that $\cos c = 0$. Define $\pi := 2c$, that is, $c = \pi/2$. Then $0 < c < 2$ implies that $0 < \pi < 4$ and this is the only number between 0 and 4 such that $\cos(\pi/2) = 0$. Since cos is strictly decreasing on $[0, 2]$, we have $\cos x > 0$ for $0 < x < \pi/2$. To see that in fact, $3 < \pi < 4$, we just need to show that $\cos(3/2) > 0$; this implies that $3/2 < c < 2$ and therefore $3 < \pi < 4$. Plugging in $x = 3/2$ into the formula for $\cos x$, we get

$$\cos\frac{3}{2} = \left(1 - \frac{3^2}{2^2\,2!}\right) + \left(\frac{3^4}{2^4\,4!} - \frac{3^6}{2^6\,6!}\right) + \left(\frac{3^8}{2^8\,8!} - \frac{3^{10}}{2^{10}\,10!}\right) + \cdots .$$

The first term is negative and equals $1 - 9/8 = -1/8$, while, as the reader can check, all the rest of the parentheses are positive numbers. In particular (after a lot of scratch work figuring out the number in the second parentheses), we obtain

$$\cos\frac{3}{2} > \left(1 - \frac{3^2}{2^2\,2!}\right) + \left(\frac{3^4}{2^4\,4!} - \frac{3^6}{2^6\,6!}\right) = -\frac{1}{8} + \frac{3^3 \cdot 37}{2^{10} \cdot 5} = \frac{359}{2^{10} \cdot 5} > 0.$$

□

The number $\pi/180$ is called a **degree**. Thus, $\pi/2 = 90 \cdot \pi/180$ is the same as 90 degrees, which we write as $90°$, $\pi = 180 \cdot \pi/180$ is the same as $180°$, etc.

4.7.3. Properties of π.

As we already stated, the approach we have taken to introduce π has been completely analytical without reference to triangles or circles, but surely the π we have defined and the π you have grown up with must be the same. We now show that the π we have defined is not an imposter, but indeed does have all the properties of the π that you have grown to love.

We first state some of the well-known trig identities involving π that you learned in high school, but now we even prove them for complex variables.

THEOREM 4.37. *The following identities hold:*

$$\cos(\pi/2) = 0, \quad \cos(\pi) = -1, \quad \cos(3\pi/2) = 0, \quad \cos(2\pi) = 1$$
$$\sin(\pi/2) = 1, \quad \sin(\pi) = 0, \quad \sin(3\pi/2) = -1, \quad \sin(2\pi) = 0.$$

Moreover, for any complex number z, we have the following addition formulas:

$$\cos\left(z + \frac{\pi}{2}\right) = -\sin z, \quad \sin\left(z + \frac{\pi}{2}\right) = \cos z,$$
$$\cos(z + \pi) = -\cos z, \quad \sin(z + \pi) = -\sin z,$$
$$\cos(z + 2\pi) = \cos z, \quad \sin(z + 2\pi) = \sin z.$$

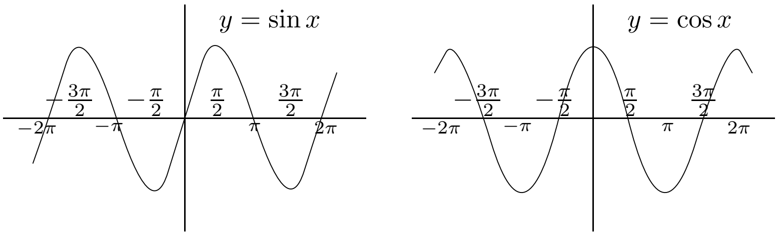


FIGURE 4.11. Our definitions of sine and cosine have the same properties as the ones you learned in high school!

PROOF. We know that $\cos(\pi/2) = 0$ and, by *(1)* of Lemma 4.35, $\sin(\pi/2) > 0$, therefore since

$$\sin^2(\pi/2) = 1 - \cos^2(\pi/2) = 1,$$

we must have $\sin(\pi/2) = 1$. The double angle formulas now imply that

$$\cos(\pi) = \cos^2(\pi/2) - \sin^2(\pi/2) = -1, \quad \sin(\pi) = 2\cos(\pi/2)\sin(\pi/2) = 0,$$

and by another application of the double angle formulas, we get

$$\cos(2\pi) = 1, \quad \sin(2\pi) = 0.$$

The facts just proved plus the addition formulas for cosine and sine in Property *(3)* of Theorem 4.34 imply the last six formulas above; for example,

$$\cos\left(z + \frac{\pi}{2}\right) = \cos z\, \cos\frac{\pi}{2} - \sin z\, \sin\frac{\pi}{2} = -\sin z,$$

and the other formulas are proved similarly. Finally, setting $z = \pi$ into

$$\cos\left(z + \frac{\pi}{2}\right) = -\sin z, \quad \sin\left(z + \frac{\pi}{2}\right) = \cos z$$

prove that $\cos(3\pi/2) = 0$ and $\sin(3\pi/2) = -1$. □

The last two formulas in Theorem 4.37 (plus an induction argument) imply that cos and sin are **periodic** (with period 2π) in the sense that for any $n \in \mathbb{Z}$,

$$\cos(z + 2\pi n) = \cos z, \quad \sin(z + 2\pi n) = \sin z. \tag{4.35}$$

Now, substituting $z = \pi$ into $e^{iz} = \cos z + i \sin z$ and using that $\cos \pi = -1$ and $\sin \pi = 0$, we get $e^{i\pi} = -1$, or by bringing -1 to the left we get perhaps most important equation in all of mathematics (at least to some mathematicians!):[10]

$$\boxed{e^{i\pi} + 1 = 0.}$$

In one shot, this single equation contains the five "most important" constants in mathematics: 0, the additive identity, 1, the multiplicative identity, i, the imaginary unit, and the constants e, the base of the exponential function, and π, the fundamental constant of geometry.

Now consider the following theorem, which essentially states that the graphs of cosine and sine go "up and down" as you think they should; see Figure 4.11.

[10] *[after proving Euler's formula $e^{i\pi} = -1$ in a lecture] Gentlemen, that is surely true, it is absolutely paradoxical; we cannot understand it, and we don't know what it means. But we have proved it, and therefore we know it is the truth. Benjamin Peirce (1809–1880). Quoted in E Kasner and J Newman* [**110**].

Theorem 4.38 (**Oscillation theorem**). *On the interval* $[0, 2\pi]$*, the following monotonicity properties of* cos *and* sin *hold:*

(1) cos *decreases from* 1 *to* -1 *on* $[0, \pi]$ *and increases from* -1 *to* 1 *on* $[\pi, 2\pi]$.
(2) sin *increases from* 0 *to* 1 *on* $[0, \pi/2]$ *and increases from* -1 *to* 0 *on* $[3\pi/2, 2\pi]$*, and decreases from* 1 *to* -1 *on* $[\pi/2, 3\pi/2]$.

Proof. From Lemma 4.35 we know that cos is strictly decreasing from 1 to 0 on $[0, \pi/2]$ and from this same lemma we know that sin is positive on $(0, \pi/2)$. Therefore by the Pythagorean identity,

$$\sin x = \sqrt{1 - \cos^2 x}$$

on $[0, \pi/2]$. Since cos is positive and strictly decreasing on $[0, \pi/2]$, this formula implies that sin is strictly increasing on $[0, \pi/2]$. Replacing z by $x - \pi/2$ in the formulas

$$\cos\left(z + \frac{\pi}{2}\right) = -\sin z, \quad \sin\left(z + \frac{\pi}{2}\right) = \cos z$$

found in Theorem 4.37, give the new formulas

$$\cos x = -\sin\left(x - \frac{\pi}{2}\right), \quad \sin x = \cos\left(x - \frac{\pi}{2}\right).$$

The first of these new formulas plus the fact that sin is increasing on $[0, \pi/2]$ show that cos is decreasing on $[\pi/2, \pi]$, while the second of these formulas plus the fact that cos is decreasing on $[0, \pi/2]$ show that sin is also decreasing on $[\pi/2, \pi]$. Finally, the formulas

$$\cos x = -\cos(x - \pi), \quad \sin x = -\sin(x - \pi),$$

also obtained as a consequence of Theorem 4.37, and the monotone properties already established for cos and sin on $[0, \pi]$, imply the rest of the monotone properties in *(1)* and *(2)* of cos and sin on $[\pi, 2\pi]$. □

In geometric terms, the following theorem states that as θ moves from 0 to 2π, the point $f(\theta) = (\cos\theta, \sin\theta)$ in $\mathbb{R}^2$ moves around the unit circle. (However, because we like complex notation, we shall write $(\cos\theta, \sin\theta)$ as the complex number $\cos\theta + i\sin\theta = e^{i\theta}$ in the theorem.)

Theorem 4.39 (**π and the unit circle**). *For a real number* θ*, define*

$$f(\theta) := e^{i\theta} = \cos\theta + i\sin\theta.$$

Then $f : \mathbb{R} \longrightarrow \mathbb{C}$ *is a continuous function and has range equal to the unit circle*

$$\mathbb{S}^1 := \{(a, b) \in \mathbb{R}^2 \,;\, a^2 + b^2 = 1\} = \{z \in \mathbb{C} \,;\, |z| = 1\}.$$

Moreover, for each $z \in \mathbb{S}^1$ *there exists a unique* θ *with* $0 \leq \theta < 2\pi$ *such that* $f(\theta) = z$*. Finally,* $f(\theta) = f(\phi)$ *if and only if* $\theta - \phi$ *is an integer multiple of* 2π.

Proof. Since the exponential function is continuous, so is the function f, and by the Pythagorean identity, $\cos^2\theta + \sin^2\theta = 1$, so we also know that f maps into the unit circle. Given z in the unit circle, we can write $z = a + ib$ where $a^2 + b^2 = 1$. We prove that there exists a unique $0 \leq \theta < 2\pi$ such that $f(\theta) = z$, that is, such that $\cos\theta = a$ and $\sin\theta = b$. Now either $b \geq 0$ or $b < 0$. Assume that $b \geq 0$; the case when $b < 0$ is proved in a similar way. Since, according to Theorem 4.38, $\sin\theta < 0$ for all $\pi < \theta < 2\pi$, and we are assuming $b \geq 0$, there is no θ with $\pi < \theta < 2\pi$ such that $f(\theta) = z$. Hence, we just have to show there is a unique $\theta \in [0, \pi]$ such that $f(\theta) = z$. Since $a^2 + b^2 = 1$, we have $-1 \leq a \leq 1$ and $0 \leq b \leq 1$. Since cos strictly decreases from 1 to -1 on $[0,\pi]$, by the intermediate value theorem there is a unique value $\theta \in [0,\pi]$ such that $\cos\theta = a$. The identity

$$\sin^2\theta = 1 - \cos^2\theta = 1 - a^2 = b^2,$$

and the fact that $\sin\theta \geq 0$, because $0 \leq \theta \leq \pi$, imply that $b = \sin\theta$.

We now prove the last assertion of our theorem. Let θ and ϕ be real numbers and suppose that $f(\theta) = f(\phi)$. Let n be the unique integer such that

$$n \leq \frac{\theta - \phi}{2\pi} < n + 1.$$

Multiplying everything by 2π and subtracting by $2\pi n$, we obtain

$$0 \leq \theta - \phi - 2\pi n < 2\pi.$$

By periodicity (see (4.35)),

$$f(\theta - \phi - 2\pi n) = f(\theta - \phi) = e^{i(\theta-\phi)} = e^{i\theta}e^{-i\phi} = f(\theta)/f(\phi) = 1.$$

Since $\theta - \phi - 2\pi n$ is in the interval $[0, 2\pi)$ and $f(0) = 1$ also, by the uniqueness we proved in the previous paragraph, we conclude that $\theta - \phi - 2\pi n = 0$. This completes the proof of the theorem. □

We now solve trigonometric equations. Notice that Property *(2)* of the following theorem shows that cos vanishes at exactly $\pi/2$ and all its π translates and *(3)* shows that sin vanishes at exactly all integer multiples of π, again, well-known facts from high school! However, we consider complex variables instead of just real variables.

THEOREM 4.40. *For complex numbers z and w,*

(1) $e^z = e^w$ *if and only if* $z = w + 2\pi i n$ *for some integer* n.
(2) $\cos z = 0$ *if and only if* $z = n\pi + \pi/2$ *for some integer* n.
(3) $\sin z = 0$ *if and only if* $z = n\pi$ *for some integer* n.

PROOF. The "if" statements follow from Theorem 4.37 so we are left to prove the "only if" statements. Suppose that $e^z = e^w$. Then $e^{z-w} = 1$. Hence, it suffices to prove that $e^z = 1$ implies that z is an integer multiple of $2\pi i$. Let $z = x + iy$ for real numbers x and y. Then,

$$1 = |e^{x+iy}| = |e^x\, e^{iy}| = e^x.$$

Since the exponential function on the real line in one-to-one, it follows that $x = 0$. Now the equation $1 = e^z = e^{iy}$ implies, by Theorem 4.39, that y must be an integer multiple of 2π. Hence, $z = x + iy = iy$ is an integer multiple of $2\pi i$.

Assume that $\sin z = 0$. Then by definition of $\sin z$, we have $e^{iz} = e^{-iz}$. By *(1)*, we have $iz = -iz + 2\pi i n$ for some integer n. Solving for z, we get $z = \pi n$. Finally, the identity

$$\sin\left(z + \frac{\pi}{2}\right) = \cos z$$

and the result already proved for sine shows that $\cos z = 0$ implies that $z = n\pi + \pi/2$ for some integer n. □

As a corollary of this theorem we see that the domain of $\tan z = \sin z/\cos z$ and $\sec z = 1/\cos z$ consists of all complex numbers except integer multiples of $\pi/2$.

4.7.4. Which is larger, π^e or e^π? Of course, one can simply check using a calculator that e^π is greater. Here's a mathematical proof following [**196**]. First recall that $1+x < e^x$ for any positive real x. Hence, as powers preserve inequalities, for any $x, y > 0$, we obtain

$$\left(1 + \frac{x}{y}\right)^y < (e^{x/y})^y = e^x.$$

In Section 3.7, we noted that $e < 3$. Since $3 < \pi$, we have $\pi - e > 0$. Now setting $x = \pi - e > 0$ and $y = e$ into the above equation, we get

$$\left(1 + \frac{\pi - e}{e}\right)^e = \left(\frac{\pi}{e}\right)^e < e^{\pi - e},$$

which, after multiplying by e^e, gives the inequality $\pi^e < e^\pi$.

By the way, speaking about e^π, Charles Hermite (1822–1901) made a fascinating discover that for many values of n, $e^{\pi\sqrt{n}}$ is an "**almost integer**" [**47**, p. 80]. For example, if you go to a calculator, you'll find that when $n = 1$, e^π is not almost an integer, but $e^\pi - \pi$ is:

$$\boxed{e^\pi - \pi \approx 20.}$$

In fact, $e^\pi - \pi = 19.999099979\ldots$. When $n = 163$, we get the incredible approximation

$$\boxed{e^{\pi\sqrt{163}} = 262537412640768743.99999999999992\ldots} \tag{4.36}$$

Check out $e^{\pi\sqrt{58}}$. Isn't it amazing how e and π show up in the strangest places?

4.7.5. Plane geometry and polar representations of complex numbers. Given a nonzero complex number z, we can write $z = r\,\omega$ where $r = |z|$ and $\omega = z/|z|$. Notice that $|\omega| = 1$, so from our knowledge of π and the unit circle (Theorem 4.39) we know that there is a unique $0 \le \theta < 2\pi$ such that

$$\frac{z}{|z|} = e^{i\theta} = \cos\theta + i\sin\theta.$$

Therefore,

$$z = re^{i\theta} = r\big(\cos\theta + i\sin\theta\big).$$

This is called the **polar representation** of z. We can relate this representation to the familiar "polar coordinates" on $\mathbb{R}^2$ as follows. Recall that $\mathbb{C}$ is really just $\mathbb{R}^2$. Let $z = x + iy$, which remember is the same as $z = (x, y)$ where $i = (0, 1)$. Then

$$r = |z| = \sqrt{x^2 + y^2}$$

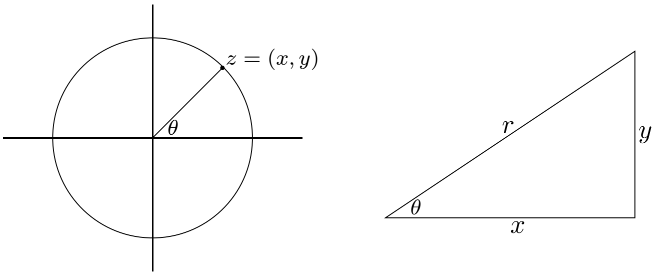
is just the familiar radial distance of (x, y) to the origin. Equating the real and imaginary parts of the equation $\cos\theta + i\sin\theta = z/|z|$, we get the two equations

$$\cos\theta = \frac{x}{r} = \frac{x}{\sqrt{x^2 + y^2}} \quad \text{and} \quad \sin\theta = \frac{y}{r} = \frac{y}{\sqrt{x^2 + y^2}}. \tag{4.37}$$

Summarizing: The equation $(x, y) = z = re^{i\theta} = r\big(\cos\theta + i\sin\theta\big)$ is equivalent to

$$x = r\,\cos\theta \quad \text{and} \quad y = r\,\sin\theta.$$

We call (r, θ) the **polar coordinates** of the point $z = (x, y)$. When z is drawn as a point in $\mathbb{R}^2$, r represents the distance of z to the origin and θ represents (or rather, is *by definition*) the **angle** that z makes with the positive real axis; see Figure 4.12. In elementary calculus, one usually studies polar coordinates without



$\cos\theta = \frac{x}{r}$

$\sin\theta = \frac{y}{r}$

FIGURE 4.12. The familiar concept of angle.

introducing complex numbers, however, we prefer the complex number approach and in particular, the single notation $z = re^{i\theta}$ instead of the pair notation $x = r\cos\theta$ and $y = r\sin\theta$. We have taken $0 \le \theta < 2\pi$, but it will be very convenient to allow θ to represent *any* real number. In this case, $z = re^{i\theta}$ is not attached to a unique choice of θ, but by our knowledge of π and the unit circle, we know that any two such θ's differ by an integer multiple of 2π. Thus, the polar coordinates (r, θ) and $(r, \theta + 2\pi n)$ represent the same point for any integer n.

Summarizing this section, we have seen that

All that you thought about trigonometry is true!

In particular, from (4.37) and adding the formula $\tan\theta = \sin\theta/\cos\theta = y/x$, from Figure 4.12 we see that

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotonus}}, \quad \sin\theta = \frac{\text{opposite}}{\text{hypotonus}}, \quad \tan\theta = \frac{\text{opposite}}{\text{adjacent}},$$

just as you learned from high school!

EXERCISES 4.7.

1. Here are some values of the trigonometric functions.
   (a) Find $\sin i$, $\cos i$, and $\tan(1+i)$ (in terms of e and i).
   (b) Using various trig identities (no triangles allowed!), prove the following well-known values of sine and cosine: $\sin(\pi/4) = \cos(\pi/4) = 1/\sqrt{2}$, $\sin(\pi/6) = \cos(\pi/3) = 1/2$, and $\sin(\pi/3) = \cos(\pi/6) = \sqrt{3}/2$.
   (c) Using trig identities, find $\sin(\pi/8)$ and $\cos(\pi/8)$.
2. In this problem we find a very close estimate of π. Prove that for $0 < x < 2$, we have
$$\cos x < 1 - \frac{x^2}{2} + \frac{x^4}{24}.$$
   Use this fact to prove that $3/2 < \pi/2 < \sqrt{6 - 2\sqrt{3}}$, which implies that $3 < \pi < 2\sqrt{6 - 2\sqrt{3}} \approx 3.185$. We'll get a much better estimate in Section 4.10.
3. Using the series representations (4.33) for $\sin z$ and $\cos z$, find the limits
$$\lim_{z\to 0} \frac{\sin z}{z}, \quad \lim_{z\to 0} \frac{\sin z - z}{z^3}, \quad \lim_{z\to 0} \frac{\cos z - 1 + z^2/2}{z^3}, \quad \lim_{z\to 0} \frac{\cos z - 1 + z^2/2}{z^4}.$$
4. Prove some of the following identities:

(a) For $z, w \in \mathbb{C}$,

$$\begin{aligned} 2\sin z \sin w &= \cos(z-w) - \cos(z+w), \\ 2\cos z \cos w &= \cos(z-w) + \cos(z+w), \\ 2\sin z \cos w &= \sin(z+w) + \sin(z-w), \\ \tan(z+w) &= \frac{\tan z + \tan w}{1 - \tan z \, \tan w}, \\ 1 + \tan^2 z &= \sec^2 z, \quad \cot^2 z + 1 = \csc^2 z. \end{aligned}$$

(b) If $x \in \mathbb{R}$, then for any natural number n,

$$\begin{aligned} \cos nx &= \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \binom{n}{2k} \cos^{n-2k} x \, \sin^{2k} x, \\ \sin nx &= \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} (-1)^k \binom{n}{2k+1} \cos^{n-2k-1} x \, \sin^{2k+1} x, \end{aligned}$$

where $\lfloor t \rfloor$ is the greatest integer less than or equal to $t \in \mathbb{R}$. Suggestion: Expand the left-hand side of de Moivre's formula using the binomial theorem.

(c) Prove that

$$\sin^2 \frac{\pi}{5} = \frac{5-\sqrt{5}}{8}, \quad \cos^2 \frac{\pi}{5} = \frac{3+\sqrt{5}}{8}, \quad \cos \frac{\pi}{5} = \frac{1+\sqrt{5}}{4}.$$

Suggestion: What if you consider $x = \pi/5$ and $n = 5$ in the equation for $\sin nx$ in Part (b)?

5. Prove that for $0 \le r < 1$ and $\theta \in \mathbb{R}$,

$$\boxed{\sum_{n=0}^{\infty} r^n \cos(n\theta) = \frac{1 - r\cos\theta}{1 - 2r\cos\theta + r^2} \quad , \quad \sum_{n=1}^{\infty} r^n \sin(n\theta) = \frac{r\sin\theta}{1 - 2r\cos\theta + r^2}.}$$

Suggestion: Let $z = re^{i\theta}$ in the geometric series $\sum_{n=0}^{\infty} z^n$.

6. Prove that if $e < \beta$, then $\beta^e < e^\beta$.

7. Here's a very neat problem posed by D.J. Newman [**156**].

(i) Prove that

$$\lim_{n\to\infty} n \, \sin(2\pi \, e \, n!) = 2\pi.$$

where e is Euler's number. Suggestion: Start by multiplying $e = 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \frac{1}{(n+1)!} + \cdots$ by $2\pi n!$ and see what happens.

(ii) Prove, using (i), that e is irrational.

8. (**Hyperbolic functions**) In this problem we study the hyperbolic functions.

(a) Show that

$$\begin{aligned} \sinh(z+w) &= \sinh z \cosh w + \cosh z \sinh w, \\ \cosh(z+w) &= \cosh z \cosh w + \sinh z \sinh w, \\ \sinh(2z) &= 2\cosh z \sinh z \quad , \quad \cosh^2 z - \sinh^2 z = 1 \end{aligned}$$

(b) If $z = x + iy$, prove that

$$\begin{aligned} \sinh z = \sinh x \cos y + i\cosh x \sin y, &\quad \cosh z = \cosh x \cos y + i \sinh x \sin y \\ |\sinh z|^2 = \sinh^2 x + \sin^2 y, &\qquad |\cosh z|^2 = \sinh^2 x + \cos^2 y. \end{aligned}$$

Determine all $z \in \mathbb{C}$ such that $\sinh z$ is real. Do the same for $\cosh z$. Determine all the zeros of $\sinh z$ and $\cosh z$.

(c) Prove that if $z = x + iy$, then

$$\sin z = \sin x \cosh y + i \cos x \sinh y, \quad \cos z = \cos x \cosh y - i \sin x \sinh y.$$

Determine all $z \in \mathbb{C}$ such that $\sin z$ is real. Do the same for $\cos z$.

9. Here is an interesting geometric problem. Let $z \in \mathbb{C}$ and let $G(n, r)$ denote a regular n-gon ($n \geq 3$) of radius r centered at the origin of $\mathbb{C}$. In this problem we find a formula for the sum of the squares of the distances from z to the vertices of $G(n, r)$. Using complex numbers, this problem is not too difficult to solve. Proceed as follows.

(i) Show that $0 = \sum_{k=1}^n e^{2\pi ik/n} = e^{2\pi i/n} + \left(e^{2\pi i/n}\right)^2 + \cdots + \left(e^{2\pi i/n}\right)^n$.

(ii) Show that

$$\sum_{k=1}^n \left| z - re^{2\pi ik/n} \right|^2 = n(|z|^2 + r^2).$$

Interpret this equation in the context of our problem.

10. In this problem we consider "Thomae-like" functions. Prove that the following functions are continuous at the irrationals and discontinuous at the rationals.

(a) Define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} \sin(1/q) & \text{if } x \in \mathbb{Q} \text{ and } x = p/q \text{ in lowest terms and } q > 0, \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

(b) Define $g : (0, \infty) \longrightarrow \mathbb{R}$ by

$$g(x) = \begin{cases} p \sin(1/q) & \text{if } x \in \mathbb{Q} \text{ and } x = p/q \text{ in lowest terms and } q > 0, \\ x & \text{if } x \text{ is irrational.} \end{cases}$$

11. In this problem we define π using only the most elementary properties of cosine and sine. (See [**209**, p. 160] for another proof). Assume that you are given continuous functions $\cos, \sin : \mathbb{R} \longrightarrow \mathbb{R}$ such that

(a) $\cos^2 x + \sin^2 x = 1$ for all $x \in \mathbb{R}$.

(b) $\cos(0) = 1$ and $\sin x$ is positive for $x > 0$ sufficiently small.

(c) $\sin(x \pm y) = \sin x \, \cos y \pm \cos x \, \sin y$ for all $x, y \in \mathbb{R}$.

Based on these three properties of cosine and sine, we shall prove that

$$\pi := 2 \cdot \inf A \ , \quad \text{where } A = \{x > 0\, ;\, \cos x = 0\},$$

is well-defined, which amounts to showing that $A \neq \varnothing$. Assume, by way of contradiction, that $A = \varnothing$. Now proceed as follows.

(i) First establish the following identity: For any $x, y \in \mathbb{R}$,

$$\sin x - \sin y = 2 \cos \frac{x+y}{2} \, \sin \frac{x-y}{2}.$$

(ii) Show that $\cos x > 0$ for all $x \geq 0$.

(iii) Using (a) show that $\sin : [0, \infty) \longrightarrow \mathbb{R}$ is strictly increasing and use this to show that $\cos : [0, \infty) \longrightarrow \mathbb{R}$ is strictly decreasing.

(iv) Show that $L := \lim_{x \to \infty} \cos x$ exists and $\lim_{x \to \infty} \sin x = \sqrt{1 - L^2}$.

(v) Prove that $\sin 2x = 2 \cos x \, \sin x$ for all $x \in \mathbb{R}$ and then prove that $L = \frac{1}{2}$.

(vi) Using the identity in (i), prove that for any $y \in \mathbb{R}$ we have $\sin y \leq 0$. This contradicts that $\sin x > 0$ for x sufficiently small.

(vii) Thus, the assumption that $A = \varnothing$ must have been false and hence π is well-defined. Now that we know π is well-defined we can use this new definition to re-prove some properties we already verified in the text. For example, prove that $\cos(\pi/2) = 0$ and from (i), show that $\sin x$ is strictly increasing on $[0, \pi/2]$. Prove that $\sin(\pi/2) = 1$ and $\cos x$ is strictly decreasing on $[0, \pi/2]$.

4.8. ★ Three proofs of the fundamental theorem of algebra (FTA)

In elementary calculus you were exposed to the "method of partial fractions" to integrate rational functions, in which you had to factor polynomials. The necessity to factor polynomials for the method of partial fractions played a large rôle in the race to prove the fundamental theorem of algebra; see [**61**] for more on this history, especially Euler's part in this theorem. It was Carl Friedrich Gauss (1777–1855) who first proved the fundamental theorem of algebra, as part of his doctoral thesis (1799) entitled "A new proof of the theorem that every integral rational algebraic function[11] can be decomposed into real factors of the first or second degree" (see e.g. [**36**, p. 499]). We present three independent and different guises of one of the more elementary and popular "topological" proofs of the theorem, except we shall work with general complex polynomials, that is, polynomials with complex coefficients.

4.8.1. Our first proof of the FTA. Our first proof is found in the article by Remmert [**186**]. This proof could have actually been presented immediately after Section 4.4, but we have chosen to save the proof till now because it fits so well with roots of complex numbers that we'll touch on in Section 4.8.3.

Given $n \in \mathbb{N}$ and $z \in \mathbb{C}$, a complex number ξ is called an ***n*-th root** of z if $\xi^n = z$. A natural question is: Does every $z \in \mathbb{C}$ have an n-th root? Notice that if $z = 0$, then $\xi = 0$ is an n-th root of z and is the only n-th root (since a nonzero number cannot be an n-th root of 0 because the product of nonzero complex numbers is nonzero). Thus, for existence purposes we may assume that z is nonzero. Now certainly if $n = 1$, then z has a one root; namely $\xi = z$. If $n = 2$ and if z is a real positive number, then we know z has a square root and if z is a real negative number, then $\xi = i\sqrt{-z}$ is a square root of z. If $z = a + ib$, where $b \neq 0$, then the numbers

$$\xi = \pm\left(\sqrt{\frac{|z| + a}{2}} + i\frac{b}{|b|}\sqrt{\frac{|z| - a}{2}}\right).$$

are square roots of z, as the reader can easily verify; see Problem 8. What about higher order roots for nonzero complex numbers? In the following lemma we prove that any complex number has an n-th root. In Subsection 4.8.3 we'll give another proof of this lemma using facts about exponential and trigonometric functions developed in the previous sections. However, the following proof is interesting because it is completely elementary in that it avoids any reference to these functions.

LEMMA 4.41. *Any complex number has an n-th root.*

PROOF. Let $z \in \mathbb{C}$, which we may assume is nonzero. We shall prove that z has an n-th root using strong induction. We already know that z has n-th roots for $n = 1, 2$. Let $n > 2$ and assume that z has roots for all natural numbers less than n, we shall prove that z has an n-th root.

Suppose first that n is even, say $n = 2m$ for some natural number $m > 2$. Then we are looking for a complex number ξ such that $\xi^{2m} = z$. By our discussion before this lemma, we know that there is a number η such that $\eta^2 = z$ and since $m < n$,

[11] In plain English, a polynomial with real coefficients. You can find a beautiful translation of Gauss' thesis by Ernest Fandreyer at http://www.fsc.edu/library/documents/Theorem.pdf. Gauss' proof was actually incorrect, but he published a correct version in 1816.

by induction hypothesis, we know there is a number ξ such that $\xi^m = \eta$. Then

$$\xi^n = \xi^{2m} = (\xi^m)^2 = \eta^2 = z,$$

and we've found an n-th root of z.

Suppose now that n is odd. If z is a nonnegative real number, then we know that z has a real n-th root, so we may assume that z is not a nonnegative real number. Choose a complex number η such that $\eta^2 = z$. Then for $x \in \mathbb{R}$, consider the polynomial $p(x)$ given by taking the imaginary part of $\eta(x-i)^n$:

$$p(x) := \operatorname{Im}\big[\eta(x-i)^n\big] = \frac{1}{2i}\big[\eta(x-i)^n - \overline{\eta}(x+i)^n\big],$$

where we used Property *(4)* of Theorem 2.43 that $\operatorname{Im} w = \frac{1}{2i}(w - \overline{w})$ for any complex number w. Expanding $(x-i)^n$ using the binomial theorem, we see that

$$p(x) = \operatorname{Im}(\eta)\, x^n + \text{ lower order terms in } x.$$

Since η is not real, the coefficient in front of x^n is nonzero, so $p(x)$ is an n-th degree polynomial in x with real coefficients. In Problem 2 of Section 4.4 we noted that all odd degree real-valued polynomials have a real root, so there is some $c \in \mathbb{R}$ with $p(c) = 0$. For this c, we have

$$\eta(c-i)^n - \overline{\eta}(c+i)^n = 0.$$

After a little manipulation, and using that $\eta^2 = z$, we get

$$\frac{(c+i)^n}{(c-i)^n} = \frac{\eta}{\overline{\eta}} = \frac{\eta^2}{|\eta|^2} = \frac{z}{|z|} \quad \Longrightarrow \quad |z|\,\frac{(c+i)^n}{(c-i)^n} = z,$$

It follows that $\xi = \sqrt[n]{|z|}\,\frac{c+i}{c-i}$ satisfies $\xi^n = z$ and our proof is now complete. □

We now present our first proof of the celebrated fundamental theorem of algebra. The following proof is a very elementary proof of Gauss' famous result in the sense that looking through the proof, we see that the nontrivial results we use are kept at a minimum:

(1) The Bolzano-Weierstrass theorem.
(2) Any nonzero complex number has a k-th root.

For other presentations of basically the same proof, see [**69**], [**222**], [**191**], or (one of my favorites) [**185**].

Theorem 4.42 (**The fundamental theorem of algebra, Proof I**). *Any complex polynomial of positive degree has at least one complex root.*

Proof. Let $p(z) = a_n\, z^n + a_{n-1}\, z^{n-1} + \cdots + a_1\, z + a_0$ be a polynomial with complex coefficients, $n \geq 1$ with $a_n \neq 0$. We prove this theorem in four steps.

Step 1: We begin by proving a simple, but important, inequality. Since

$$|p(z)| = |a_n\, z^n + \cdots + a_0| = |z|^n \left| a_n + \frac{a_{n-1}}{z} + \frac{a_{n-2}}{z^2} + \cdots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right|,$$

for $|z|$ sufficiently large the absolute value of the sum of all the terms to the right of a_n can be made less than, say $|a_n|/2$. Therefore,

$$|p(z)| \geq \frac{|a_n|}{2} \cdot |z|^n\,, \quad \text{for } |z| \text{ sufficiently large.} \tag{4.38}$$

Step 2: We now prove that there exists a point $c \in \mathbb{C}$ such that $|p(c)| \leq |p(z)|$ for all $z \in \mathbb{C}$. The proof of this involves the Bolzano-Weierstrass theorem. Define

$$m := \inf A, \qquad A := \{|p(z)|\,;\, z \in \mathbb{C}\}.$$

This infimum certainly exists since A is nonempty and bounded below by zero. Since m is the greatest lower bound of A, for each $k \in \mathbb{N}$, $m + 1/k$ is no longer a lower bound, so there is a point $z_k \in \mathbb{C}$ such that $m \leq |p(z_k)| < m+1/k$. By (4.38), the sequence $\{z_k\}$ must be bounded, so by the Bolzano-Weierstrass theorem, this sequence has a convergent subsequence $\{w_k\}$. If c is the limit of this subsequence, then by continuity of polynomials, $|p(w_k)| \to |p(c)|$ and since $m \leq |p(z_k)| < m+1/k$ for all k, by the squeeze theorem we must have $|p(c)| = m$.

Step 3: The rest of the proof involves showing that the minimum m must be zero, which shows that $p(c) = 0$, and so c is a root of $p(z)$. To do so, we introduce an auxiliary polynomial $q(z)$ as follows. Let us suppose, for sake of contradiction, that $p(c) \neq 0$. Define $q(z) := p(z + c)/p(c)$. Then $|q(z)|$ has a minimum at the point $z = 0$, the minimum being $|q(0)| = |1| = 1$. Since $q(0) = 1$, we can write

$$q(z) = b_n\, z^n + \cdots + 1 = b_n\, z^n + \cdots + b_k z^k + 1, \tag{4.39}$$

where k is the smallest natural number such that $b_k \neq 0$. In our next step we shall prove that 1 is in fact not the minimum of $|q(z)|$, which gives a contradiction.

Step 4: By our lemma, $-1/b_k$ has a k-th root a, so that $a^k = -1/b_k$. Then $|q(az)|$ also has a minimum at $z = 0$, and

$$q(bz) = 1 + b_k(az)^k + \cdots = 1 - z^k + \cdots\,,$$

where $\cdots$ represents terms of higher degree than k. Thus, we can write

$$q(az) = 1 - z^k + z^{k+1} r(z),$$

where $r(z)$ is a polynomial of degree at most $n - (k+1)$. Let $z = x$, a real number with $0 < x < 1$, be so small that $x\,|r(x)| < 1$. Then,

$$\begin{aligned}|q(ax)| = |1 - x^k + x^{k+1} r(x)| &\leq |1 - x^k| + x^{k+1}|r(x)| \\ &< 1 - x^k + x^k \cdot 1 = 1 = |q(0)| \quad \Longrightarrow \quad |q(ax)| < |q(0)|.\end{aligned}$$

This shows that $|q(z)|$ does not achieve a minimum at $z = 0$, contrary to what we said earlier. Hence our assumption that $p(c) \neq 0$ must have been false and our proof is complete. □

We remark that the other two proofs of the FTA in this section (basically) only differ from this proof at the first line in **Step 4**, in how we claim that there is a complex number a with $a^k = -1/b_k$.

As a consequence of the fundamental theorem of algebra, we can prove the well-known fact that a polynomial can be factored. Let $p(z)$ be a polynomial of positive degree n with complex coefficients and let c_1 be a root of p, which we know exists by the FTA. Then from Lemma 2.52, we can write

$$p(z) = (z - c_1)\, q_1(z),$$

where $q_1(z)$ is a polynomial of degree $n - 1$ in both z and c_1. By the FTA, q_1 has a root, call it c_2. Then from Lemma 2.52, we can write $q_1(z) = (z - c_2)\, q_2(z)$ where q_2 has degree $n - 2$ and substituting q_1 into the formula for p, we obtain

$$p(z) = (z - c_1)(z - c_2)\, q_2(z).$$

Proceeding a total of $n-2$ more times in this fashion we eventually arrive at

$$p(z) = (z - c_1)(z - c_2)\cdots(z - c_n)\, q_n,$$

where q_n is a polynomial of degree zero, that is, a necessarily nonzero constant. It follows that $c_1, \ldots, c_n$ are roots of $p(z)$. Moreover, these numbers are the only roots, for if

$$0 = p(c) = (c - c_1)(c - c_2)\cdots(c - c_n)\, q_n,$$

then c must equal one of the c_k's since a product of complex numbers is zero if and only if one of the factors is zero. Summarizing, we have proved the following.

Corollary 4.43. *If $p(z)$ is a polynomial of positive degree n, then p has exactly n complex roots $c_1, \ldots, c_n$ counting multiplicities and we can write*

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

4.8.2. Our second proof of the FTA. Our second proof of the FTA is almost exactly the same as the first, but at the beginning of **Step 4** in the above proof, we use a neat trick by Searcóid [**167**] that avoids the fact that *every* complex number has an n-th root. His trick is the following lemma.

Lemma 4.44. *Let ℓ be an odd natural number, $\zeta = (1+i)/\sqrt{2}$, and let α be a complex number of length 1. Then there is a natural number ν such that*

$$|1 + \alpha\, \zeta^{2\nu\ell}| < 1.$$

Proof. Observe that

$$\zeta^2 = \frac{(1+i)(1+i)}{2} = \frac{1 + 2i + i^2}{2} = i.$$

Therefore, $1+\alpha\,\zeta^{2\nu\ell}$ simplifies to $1+\alpha\, i^{\nu\ell}$, and we shall use this latter expression for the rest of the proof. Since ℓ is odd we can write $\ell = 2m+1$ for some $m = 0, 1, 2, \ldots$, thus for any natural number ν,

$$i^{\nu\ell} = i^{\nu(2m+1)} = i^{2\nu m} \cdot i^{\nu} = (-1)^{\nu m} \cdot i^{\nu} = \begin{cases} i^{\nu} & \text{if } m \text{ is even} \\ (-i)^{\nu} & \text{if } m \text{ is odd.} \end{cases}$$

Using this formula one can check that $\{i^{\nu\ell}\,;\, \nu \in \mathbb{N}\} = \{1, i, -1, -i\}$. Observe that

$$|1 + \alpha\, i^{\nu\ell}|^2 = (1 + \alpha\, i^{\nu\ell})(1 + \overline{\alpha\, i^{\nu\ell}}) = 1 + \alpha\, i^{\nu\ell} + \overline{\alpha\, i^{\nu\ell}} + |\alpha\, i^{\nu\ell}|^2 = 2 + 2\,\mathrm{Re}(\alpha\, i^{\nu\ell}),$$

where we used that $|\alpha\, i^{\nu\ell}| = |\alpha| = 1$ and that $2\,\mathrm{Re}\, w = w + \overline{w}$ for any complex number w from Property *(4)* of Theorem 2.43. Let $\alpha = a + ib$. Then considering the various cases $i^{\nu\ell} = 1, i, -1, -i$, we get

$$\{\alpha\, i^{\nu\ell}\,;\, \nu \in \mathbb{N}\} = \{a i^{\nu\ell} + i b i^{\nu\ell}\,;\, \nu \in \mathbb{N}\} = \{a + ib,\ -b + ia,\ -a - ib,\ b - ia\}.$$

Hence, in view of the formula $|1 + \alpha\, i^{\nu\ell}|^2 = 2 + 2\,\mathrm{Re}(\alpha\, i^{\nu\ell})$, we obtain

$$\{\, |1 + \alpha\, i^{\nu\ell}|^2\,;\, \nu \in \mathbb{N}\,\} = \{2 + 2a,\ 2 - 2b,\ 2 - 2a,\ 2 + 2b\}. \tag{4.40}$$

Since $|\alpha|^2 = a^2 + b^2 = 1$, $|a| \geq 1/\sqrt{2}$ or $|b| \geq 1/\sqrt{2}$ (for otherwise $a^2 + b^2 < 1$). Let us take the case when $|a| \geq 1/\sqrt{2}$; the other case is handled similarly. If $a \geq 1/\sqrt{2}$, then $2 - 2a \leq 2 - 2/\sqrt{2} = 2 - \sqrt{2} < 1$ and a ν corresponding to $2 - 2a$ in (4.40) satisfies the conditions of this lemma. If $a \leq -1/\sqrt{2}$, then $2 + 2a \leq 2 - \sqrt{2}$, so a ν corresponding to $2 + 2a$ in (4.40) satisfies the conditions of this lemma. □

THEOREM 4.45 (**The fundamental theorem of algebra, Proof II**). *Any complex polynomial of positive degree has at least one complex root.*

PROOF. We proceed by strong induction. Certainly the FTA holds for all polynomials of first degree, therefore assume that $p(z)$ is a polynomial of degree $n \geq 2$ and suppose the FTA holds for all polynomials of degree less than n.

Now we proceed, *without changing a single word*, exactly as in Proof I up to **Step 4**, where we use the following argument in place.

Step 4 modified: Recall that the polynomial $q(z)$ in (4.39),

$$q(z) = b_n\, z^n + \cdots + b_k z^k + 1,$$

has the property that $|q(z)|$ has the minimum value 1. We claim that the k in this expression cannot equal n. To see this, for sake of contradiction, let us suppose that $k = n$. Then $q(z) = b_n\, z^n + 1$ and $q(z)$ has the property that

$$|q(z)| = |1 + b_n\, z^n| = \left|1 + \frac{b_n}{|b_n|}\, |b_n|\, z^n\right| = |1 + \alpha\, w^n|$$

has the minimum value 1, where $\alpha = b_n/|b_n|$ has unit length and $w = |b_n|^{1/n}\, z$. We derive a contradiction in three cases: $n > 2$ is even, $n = 2$, and n is odd. If $n > 2$ is even, then we can write $n = 2m$ for a natural number m with $2 \leq m < n$. By our induction hypothesis (the FTA holds for all polynomials of degree less than n), there is a number η such that $\eta^m + 1/\alpha = 0$, and, there is a number ξ such that $\xi^2 - \eta = 0$. Then

$$\xi^n = \xi^{2m} = (\xi^2)^m = \eta^m = -1/\alpha.$$

Thus, for $w = \xi$, we obtain $|1+\alpha\, w^n| = 0$, which contradicts the fact that $|1+\alpha\, w^n|$ is never less than 1. Now suppose that $n = 2$. Then by our lemma with $\ell = 1$, there is a ν such that

$$|1 + \alpha\, \zeta^{2\nu}| < 1,$$

where $\zeta = (1 + i)/\sqrt{2}$. This shows that $w = \zeta^\nu$ satisfies $|1 + \alpha\, w^n| < 1$, again contradicting the fact that $|1 + \alpha\, w^n|$ is never less than 1. Finally, suppose that $n = \ell$ is odd. Then by our lemma, there is a ν such that

$$|1 + \alpha\, \zeta^{2\nu n}| < 1,$$

where $\zeta = (1 + i)/\sqrt{2}$, which shows that $w = \zeta^{2\nu}$ satisfies $|1 + \alpha\, w^n| < 1$, again resulting in a contradiction. Therefore, $k < n$.

Now that we've proved $k < n$, we can use our induction hypothesis to conclude that there is a complex number a such that $a^k + 1/b_k = 0$, that is, $a^k = -1/b_k$. We can now proceed exactly as in **Step 4** of Proof I to finish the proof. □

4.8.3. Roots of complex numbers. Back in Section 2.7 we learned how to find n-th roots of nonnegative real numbers; we now generalize this to *complex* numbers using the polar representation of complex numbers studied in Section 4.7.

Let $n \in \mathbb{N}$ and let w be any complex number. We shall find all n-th roots of z using trigonometry. If $z = 0$, then the only ξ that works is $\xi = 0$ since the product of nonzero complex numbers is nonzero, therefore we henceforth assume that $z \neq 0$. We can write $z = re^{i\theta}$ where $r > 0$ and $\theta \in \mathbb{R}$ and given any nonzero complex ξ we can write $\xi = \rho e^{i\phi}$ where $r > 0$ and $\phi \in \mathbb{R}$. Then $\xi^n = z$ if and only if

$$\rho^n\, e^{in\phi} = r\, e^{i\theta}.$$

Taking the absolute value of both sides, and using that $|e^{in\phi}| = 1 = |e^{i\theta}|$, we get $\rho^n = r$, or $\rho = \sqrt[n]{r}$. Now cancelling off $\rho^n = r$, we see that

$$e^{in\phi} = e^{i\theta},$$

which holds if and only if $n\phi = \theta + 2\pi m$ for some integer m, or

$$\phi = \frac{\theta}{n} + \frac{2\pi m}{n}\ , \quad m \in \mathbb{Z}.$$

As the reader can easily check, any number of this form differs by an integer multiple of 2π from one of the following numbers:

$$\frac{\theta}{n},\ \frac{\theta}{n} + \frac{2\pi}{n},\ \frac{\theta}{n} + \frac{4\pi}{n}, \ldots, \frac{\theta}{n} + \frac{2\pi}{n}(n-1).$$

None of these numbers differ by an integer multiple of 2π, therefore by our knowledge of π and the unit circle, all the n numbers

$$e^{i\frac{1}{n}\left(\theta + 2\pi k\right)}, \qquad k = 0, 1, 2, \ldots, n-1$$

are distinct. Thus, there are a total of n solutions ξ to the equation $\xi^n = z$, all of them given in the following theorem.

Theorem 4.46 (**Existence of complex n-th roots**). *There are exactly n n-th roots of any nonzero complex number $z = re^{i\theta}$; the complete set of roots is given by*

$$\sqrt[n]{r}\, e^{i\frac{1}{n}\left(\theta + 2\pi k\right)} = \sqrt[n]{r}\left[\cos\frac{1}{n}\big(\theta + 2\pi k\big) + i \sin\frac{1}{n}\big(\theta + 2\pi k\big)\right], \quad k = 0, 1, 2, \ldots, n-1.$$

There is a very convenient way to write these n-th roots as we now describe. First of all, notice that

$$\sqrt[n]{r}\, e^{i\frac{1}{n}\left(\theta + 2\pi k\right)} = \sqrt[n]{r}\, e^{i\frac{\theta}{n}} \cdot e^{i\frac{2\pi k}{n}} = \sqrt[n]{r}\, e^{i\frac{\theta}{n}} \cdot \left(e^{i\frac{2\pi}{n}}\right)^k.$$

Therefore, the n-th roots of z are given by

$$\sqrt[n]{r}\, e^{i\frac{\theta}{n}} \cdot \omega^k, \qquad k = 0, 1, \ldots, n-1, \quad \text{where } \omega = e^{i\frac{2\pi}{n}} = \cos\frac{2\pi}{n} + i \sin\frac{2\pi}{n}.$$

Of all the n distinct roots, there is one called the **principal n-th root**, denoted by $\sqrt[n]{z}$, and is the n-th root given by choosing θ to satisfy $-\pi < \theta \le \pi$; thus,

$$\boxed{\sqrt[n]{z} := \sqrt[n]{r}\, e^{i\frac{\theta}{n}}\ , \quad \text{where } -\pi < \theta \le \pi.}$$

Note that if $z = x > 0$ is a positive real number, then $x = re^{i0}$ with $r = x$ and $-\pi < 0 \le \pi$, so the principal n-th root of x is just $\sqrt[n]{x}e^{i0/n} = \sqrt[n]{x}$, the usual real n-th root of x. Thus, there is no ambiguity in notation between the complex principal n-th root of a positive real number and its real n-th root.

We now give some examples.

Example 4.34. For our first example, we find the square roots of -1. Since $-1 = e^{i\pi}$, because $\cos\pi + i\sin\pi = -1 + i0$, the square roots of -1 are $e^{i(1/2)\pi}$ and $e^{i(1/2)(\pi + 2\pi)} = e^{i3\pi/2}$. Writing these numbers in terms of sine and cosine, we get i and $-i$ as the square roots of -1. Note that the principal square root of -1 is i and so $\sqrt{-1} = i$, just as we learned in high school!

Example 4.35. Next let us compute the n-th roots of unity, that is, 1. Since $1 = 1\, e^{i0}$, all the n n-th roots of 1 are given by

$$1, \omega, \omega^2, \dots, \omega^{n-1}, \quad \text{where } \omega := e^{i\frac{2\pi}{n}} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

Consider $n = 4$. In this case, $\cos\frac{2\pi}{4} + i\sin\frac{2\pi}{4} = i$, $i^2 = -1$, and $i^3 = -i$, therefore the fourth roots of unity are

$$1, i, -1, -i.$$

Since

$$\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2},$$

the cube roots of unity are

$$1, \ -\frac{1}{2} + i\frac{\sqrt{3}}{2}, \ -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

4.8.4. Our third proof of the FTA. We are now ready to prove our third proof of the FTA.

Theorem 4.47 (**The fundamental theorem of algebra, Proof III**). *Any complex polynomial of positive degree has at least one complex root.*

Proof. We proceed, *without changing a single word*, exactly as in Proof I up to **Step 4**, where we use the following in place.

Step 4 modified: At the beginning of **Step 4** in Proof I, we used Lemma 4.41 to conclude that there is a complex a such that $a^k = -1/b_k$. Now we can simply invoke Theorem 4.46 to verify that there is such a number a. Explicitly, we can just write $-1/b_k = re^{i\theta}$ and simply define $a = r^{1/k}e^{i\theta/k}$. In any case, now that we have such an a, we can proceed exactly as in **Step 4** of Proof I to finish the proof. □

Exercises 4.8.

1. Let $p(z)$ and $q(z)$ be polynomials of degree at most n.
   (a) If p vanishes at $n + 1$ distinct complex numbers, prove that $p = 0$, the zero polynomial.
   (b) If p and q agree at $n + 1$ distinct complex numbers, prove that $p = q$.
   (c) If $c_1, \dots, c_n$ (with each root repeated according to multiplicity) are roots of $p(z)$, a polynomial of degree n, prove that $p(z) = a_n(z - c_1)(z - c_2)\cdots(z - c_n)$ where a_n is the coefficient of z^n in the expression for $p(z)$.
2. Find the following roots and state which of the roots represents the principal root.
   (a) Find the cube roots of -1.
   (b) Find the square roots of i.
   (c) Find the cube roots of i.
   (d) Find the square roots of $\sqrt{3} + 3i$.
3. Geometrically (not rigorously) demonstrate that the n-th roots, with $n \geq 3$, of a nonzero complex number z are the vertices of a regular polygon.
4. Let $n \in \mathbb{N}$ and let $\omega = e^{i\frac{2\pi}{n}}$. If k is any integer that is not a multiple of n, prove that
$$1 + \omega^k + \omega^{2k} + \omega^{3k} + \cdots + \omega^{(n-1)k} = 0.$$
5. Prove by "completing the square" that any quadratic polynomial $z^2 + bz + c = 0$ with complex coefficients has two complex roots, counting multiplicities, given by
$$z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$
where $\sqrt{b^2 - 4ac}$ is the principal square root of $b^2 - 4ac$.

6. We show how the ingenious mathematicians of the past solved the general cubic equation $z^3 + bz^2 + cz + d = 0$ with complex coefficients; for the history, see [**88**].
   (i) First, replacing z with $z - b/3$, show that our cubic equation transforms into an equation of the form $z^3 + \alpha z + \beta = 0$ where α and β are complex. Thus, we may focus our attention on the equation $z^3 + \alpha\, z + \beta = 0$.
   (ii) Second, show that the substitution $z = w - \alpha/(3w)$ gives an equation of the form
   $$27(w^3)^2 + 27\beta(w^3) - \alpha^3 = 0,$$
   a quadratic equation in w^3. We can solve this equation for w^3 by the previous problem, therefore we can solve for w, and therefore we can get $z = w - \alpha/(3w)$.
   (iii) Using the technique outlined above, solve the equation $z^3 - 12z - 3 = 0$.
7. A nice application of the previous problem is finding $\sin(\pi/9)$ and $\cos(\pi/9)$.
   (i) Use de Moivre's formula to prove that
   $$\cos 3x = \cos^3 x - 3\cos x\, \sin^2 x, \qquad \sin 3x = 3\cos^2 x \sin x - \sin^3 x.$$
   (ii) Choose one of these equations and using $\cos^2 x + \sin^2 x = 1$, turn the right-hand side into a cubic polynomial in $\cos x$ or $\sin x$.
   (iii) Using the equation you get, determine $\sin(\pi/9)$ and $\cos(\pi/9)$.
8. This problem is for the classic mathematicians at heart: We find square roots without using the technology of trigonometric functions.
   (i) Let $z = a + ib$ be a nonzero complex number with $b \neq 0$. Show that $\xi = x + iy$ satisfies $\xi^2 = z$ if and only if $x^2 - y^2 = a$ and $2xy = b$.
   (ii) Prove that $x^2+y^2 = \sqrt{a^2 + b^2} = |z|$, and then $x^2 = \frac{1}{2}\big(|z|+a\big)$ and $y^2 = \frac{1}{2}\big(|z|-a\big)$.
   (iii) Finally, deduce that z must equal
   $$\xi = \pm\left(\sqrt{\frac{|z|+a}{2}} + i\frac{b}{|b|}\sqrt{\frac{|z|-a}{2}}\right).$$
9. Prove that if r is a root of a polynomial $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$, then $|r| \leq \max\big\{1, \sum_{k=0}^{n-1} |a_k|\big\}$.
10. (**Continuous dependence of roots**) Following Uherka and Sergott [**227**], we prove the following useful theorem. Let z_0 be a root of multiplicity m of a polynomial $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$. Then given any $\varepsilon > 0$, there is a $\delta > 0$ such that if $q(z) = z^n + b_{n-1}z^{n-1} + \cdots + b_0$ satisfies $|b_j - a_j| < \delta$ for all $j = 0, \ldots, n-1$, then $q(z)$ has at least m roots within ε of z_0. You may proceed as follows.
    (i) Suppose the theorem is false. Prove there is an $\varepsilon > 0$ and a sequence $\{q_k\}$ of polynomials $q_k(z) = z^n + b_{k,n-1}z^{n-1} + \cdots + b_{k,0}$ such that q_k has at most $m - 1$ roots within ε of z_0 and for each $j = 0, \ldots, n-1$, we have $b_{k,j} \to a_j$ as $k \to \infty$.
    (ii) Let $r_{k,1}, \ldots, r_{k,n}$ be the n roots of q_k. Let $R_k = (r_{k,1}, \ldots, r_{k,n}) \in \mathbb{C}^n = \mathbb{R}^{2n}$. Prove that the sequence $\{R_k\}$ has a convergent subsequence. Suggestion: Problem 9 is helpful.
    (iii) By relabelling the subsequence if necessary, we assume that $\{R_k\}$ itself converges; say $R_k = (r_{k,1}, \ldots, r_{k,n}) \to (r_1, \ldots, r_n)$. Prove that at most $m - 1$ of the r_j's can equal z_0.
    (iv) From Problem 2 in Exercises 2.10, $q_k(z) = (z - r_{k,1})(z - r_{k,2}) \cdots (z - r_{k,n})$. Prove that for each $z \in \mathbb{C}$, $\lim_{k\to\infty} q_k(z) = (z-r_1)(z-r_2)\cdots(z-r_n)$. On the other hand, using that $b_{k,j} \to a_j$ as $k \to \infty$, prove that for each $z \in \mathbb{C}$, $\lim_{k\to\infty} q_k(z) = p(z)$. Derive a contradiction.

4.9. The inverse trigonometric functions and the complex logarithm

In this section we study the inverse trigonometric functions you learned in elementary calculus. We then use these functions to derive properties of the polar angle, also called the argument of complex number. In Section 4.6 we developed the properties of real logarithms and using the logarithm we defined complex powers of positive bases. In our current section we shall extend logarithms to include complex logarithms, which are then used to define complex powers with complex bases. Finally, we use the complex logarithm to define complex inverse trigonometric functions.

4.9.1. The real-valued inverse trigonometric functions. By the oscillation theorem 4.38, we know that

$$\sin : [-\pi/2, \pi/2] \longrightarrow [-1, 1] \qquad \text{and} \qquad \cos : [0, \pi] \longrightarrow [-1, 1]$$

are both strictly monotone bijective continuous functions, sin being strictly increasing and cos being strictly decreasing. In particular, by the monotone inverse theorem, each of these functions has a strictly monotone inverse which we denote by

$$\arcsin : [-1, 1] \longrightarrow [-\pi/2, \pi/2] \qquad \text{and} \qquad \arccos : [-1, 1] \longrightarrow [0, \pi],$$

called the **inverse, or arc, sine**, which is strictly increasing, and **inverse, or arc, cosine**, which is strictly decreasing. Being inverse functions, these functions satisfy

$$\sin(\arcsin x) = x,\ -1 \le x \le 1 \qquad \text{and} \quad \arcsin(\sin x) = x, \qquad -\pi/2 \le x \le \pi/2$$

and

$$\cos(\arccos x) = x,\ -1 \le x \le 1 \qquad \text{and} \qquad \arccos(\cos x) = x,\ 0 \le x \le \pi.$$

If $0 \le \theta \le \pi$, then $-\pi/2 \le \pi/2 - \theta \le \pi/2$, so letting x denote both sides of the identity

$$\cos\theta = \sin\left(\frac{\pi}{2} - \theta\right),$$

we get $\theta = \cos^{-1} x$ and $\pi/2 - \theta = \sin^{-1} x$, which further imply that

$$\arccos x = \frac{\pi}{2} - \arcsin x, \qquad \text{for all } -1 \le x \le 1. \tag{4.41}$$

We now introduce the inverse tangent function. We first claim that

$$\tan : (-\pi/2, \pi/2) \longrightarrow \mathbb{R}$$

is a strictly increasing bijection. Indeed, since

$$\tan x = \frac{\sin x}{\cos x}$$

and sin is strictly increasing on $[0, \pi/2]$ from 0 to 1 and cos is strictly decreasing on $[0, \pi/2]$ from 1 to 0, we see that tan is strictly increasing on $[0, \pi/2)$ from 0 to ∞. Using the properties of sin and cos on $[-\pi/2, 0]$, in a similar manner one can show that tan is is strictly decreasing on $(-\pi/2, 0)$ from $-\infty$ to 0. This proves that $\tan : (-\pi/2, \pi/2) \longrightarrow \mathbb{R}$ is a strictly increasing bijection. Therefore, this function has a strictly increasing inverse, which we denote by

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

and called the **inverse, or arc, tangent**.

4.9.2. The argument of a complex number. Given a nonzero complex number z we know that we can write $z = |z|e^{i\theta}$ for some $\theta \in \mathbb{R}$ and all such θ's satisfying this equation differ by integer multiples of 2π. Geometrically, θ is interpreted as the angle z makes with the positive real axis when z is drawn as a point in $\mathbb{R}^2$. Any such angle θ is called an **argument** of z and is denoted by $\arg z$. Thus, we can write

$$z = |z|\, e^{i \arg z}.$$

We remark that $\arg z$ is not a function but is referred to as a "multiple-valued function" since $\arg z$ does not represent a single value of θ; however, any two choices for $\arg z$ differ by an integer multiple of 2π. If w is another nonzero complex number, written as $w = |w|\, e^{i\phi}$, so that $\arg w = \phi$, then

$$zw = \big(|z|\, e^{i\theta}\big)\big(|w|\, e^{i\phi}\big) = |z|\, |w|\, e^{i(\theta+\phi)},$$

which implies that

$$\arg(zw) = \arg z + \arg w.$$

We interpret this as saying that any choices for these three arguments satisfy this equation up to an integer multiple of 2π. Thus, the *argument of a product is the sum of the arguments.* What other function do you know of that takes products into sums? The logarithm of course — we shall shortly show how arg is involved in the definition of complex logarithms. Similarly, properly interpreted we have

$$\arg\left(\frac{z}{w}\right) = \arg z - \arg w.$$

With all the ambiguity in arg, mathematically it would be nice to turn arg into a function. To do so, note that given a nonzero complex number z, there is exactly one argument satisfying $-\pi < \arg z \leq \pi$; this particular angle is called the **principal argument** of z and is denoted by $\operatorname{Arg} z$. Thus, $\operatorname{Arg} : \mathbb{C} \setminus \{0\} \longrightarrow \mathbb{R}$ is characterized by the following properties:

$$\boxed{z = |z|e^{i \operatorname{Arg} z}, \qquad -\pi < \operatorname{Arg} z \leq \pi.}$$

Then all arguments of z differ from the principal one by multiples of 2π:

$$\arg z = \operatorname{Arg} z + 2\pi\, n, \qquad n \in \mathbb{Z}.$$

We can find many different formulas for $\operatorname{Arg} z$ using the inverse trig functions as follows. Writing z in terms of its real and imaginary parts: $z = x + iy$, and equating this with $|z|e^{i \operatorname{Arg} z} = |z|\cos(\operatorname{Arg} z) + i|z|\sin(\operatorname{Arg} z)$, we see that

$$\cos \operatorname{Arg} z = \frac{x}{\sqrt{x^2+y^2}} \quad \text{and} \quad \sin \operatorname{Arg} z = \frac{y}{\sqrt{x^2+y^2}}. \tag{4.42}$$

By the properties of cosine, we see that

$$-\frac{\pi}{2} < \operatorname{Arg} z < \frac{\pi}{2} \quad \Longleftrightarrow \quad x > 0.$$

Since arcsin is the inverse of sin with angles in $(-\pi/2, \pi/2)$, it follows that

$$\operatorname{Arg} z = \arcsin\left(\frac{y}{\sqrt{x^2+y^2}}\right), \qquad x > 0.$$

Perhaps the most common formula for $\operatorname{Arg} z$ when $x > 0$ is in terms of arctangent, which is derived by dividing the formulas in (4.42) to get $\tan \operatorname{Arg} z = y/x$ and then taking arctan of both sides:

$$\operatorname{Arg} z = \arctan \frac{y}{x}, \qquad x > 0.$$

We now derive a formula for $\operatorname{Arg} z$ when $y > 0$. By the properties of sine, we see that

$$0 \le \operatorname{Arg} z \le \pi \iff y \ge 0 \qquad \text{and} \qquad -\pi < \operatorname{Arg} z < 0 \iff y < 0.$$

Assuming that $y \ge 0$, that is, $0 \le \operatorname{Arg} z \le \pi$, we can take the arccos of both sides of the first equation in (4.42) to get

$$\operatorname{Arg} z = \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), \qquad y \ge 0.$$

Assume that $y < 0$, that is, $-\pi < \operatorname{Arg} z < 0$. Then $0 < -\operatorname{Arg} z < \pi$ and since $\cos \operatorname{Arg} z = \cos(-\operatorname{Arg} z)$, we get $\cos(-\operatorname{Arg} z) = x/\sqrt{x^2+y^2}$. Taking the arccos of both sides, we get

$$\operatorname{Arg} z = -\arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), \qquad y \le 0.$$

Putting together our expressions for $\operatorname{Arg} z$, we obtain the following formulas for the principal argument:

$$\boxed{\operatorname{Arg} z = \arctan \frac{y}{x} \qquad \text{if } x > 0,} \tag{4.43}$$

and

$$\boxed{\operatorname{Arg} z = \begin{cases} \arccos\left(\dfrac{x}{\sqrt{x^2+y^2}}\right) & \text{if } y \ge 0, \\ -\arccos\left(\dfrac{x}{\sqrt{x^2+y^2}}\right) & \text{if } y < 0. \end{cases}} \tag{4.44}$$

Using these formulas, we can easily prove the following theorem.

Theorem 4.48. $\operatorname{Arg} : \mathbb{C} \setminus \{0\} \longrightarrow (-\pi, \pi]$ *is continuous.*

Proof. Since

$$\mathbb{C} \setminus (-\infty, 0] = \{x + iy\,;\, x > 0\} \cup \{x + iy\,;\, y > 0\} \cup \{x + iy\,;\, y < 0\},$$

all we have to do is prove that Arg is continuous on each of these three sets. But this is easy: The formula (4.43) shows that Arg is continuous when $x > 0$, the first formula in (4.44) shows that Arg is continuous when $y > 0$, and the second formula in (4.44) shows that Arg is continuous when $y < 0$. □

4.9.3. The complex logarithm and powers. Recall from Section 4.6.2 that if $a \in \mathbb{R}$ and $a > 0$, then a real number ξ having the property that

$$e^{\xi} = a$$

is called the logarithm of a; we know that ξ always exists and is unique since $\exp : \mathbb{R} \longrightarrow (0, \infty)$ is a bijection. Of course, $\xi = \log a$ by definition of log. We now consider *complex* logarithms. We define such logarithms in an analogous way: If $z \in \mathbb{C}$ and $z \neq 0$, then a complex number ξ having the property that

$$e^{\xi} = z$$

is called a **complex logarithm** of z. The reason we assume $z \neq 0$ is that there is no complex ξ such that $e^{\xi} = 0$. We now show that nonzero complex numbers always have logarithms; however, in contrast to the case of real numbers, complex numbers have infinitely many distinct logarithms!

THEOREM 4.49. *The complex logarithms of any given nonzero complex number z are all of the form*

$$\xi = \log|z| + i\big(\operatorname{Arg} z + 2\pi\, n\big), \qquad n \in \mathbb{Z}. \tag{4.45}$$

Therefore, all complex logarithms of z have exactly the same real part $\log|z|$, but have imaginary parts that differ by integer multiples of 2π from $\operatorname{Arg} z$.

PROOF. The idea behind this proof is very simple: We write

$$z = |z| \cdot e^{i \arg z} = e^{\log|z|} \cdot e^{i \arg z} = e^{\log|z| + i \arg z}.$$

Since any argument of z is of the form $\operatorname{Arg} z + 2\pi n$ for $n \in \mathbb{Z}$, this equation shows that all the numbers in (4.45) are indeed logarithms. On the other hand, if ξ is any logarithm of z, then

$$e^{\xi} = z = e^{\log|z| + i \operatorname{Arg} z}.$$

By Theorem 4.40 we must have $\xi = \log|z| + i \operatorname{Arg} z + 2\pi i\, n$ for some $n \in \mathbb{Z}$. This completes our proof. □

To isolate one of these infinitely many logarithms we define the so-called "principal" one. For any nonzero complex number z, we define the **principal (branch of the) logarithm** of z by

$$\boxed{\operatorname{Log} z := \log|z| + i \operatorname{Arg} z.}$$

By Theorem 4.49, *all* logarithms of z are of the form

$$\operatorname{Log} z + 2\pi i\, n, \qquad n \in \mathbb{Z}.$$

Note that if $x \in \mathbb{R}$, then $\operatorname{Arg} x = 0$, therefore

$$\operatorname{Log} x = \log x,$$

our usual logarithm, so Log is an extension of the real log to complex numbers.

Example 4.36. Observe that since $\operatorname{Arg}(-1) = \pi$ and $\operatorname{Arg} i = \pi/2$ and $\log|-1| = 0 = \log|i|$, since both equal $\log 1$, we have

$$\operatorname{Log}(-1) = i\pi, \qquad \operatorname{Log} i = i\frac{\pi}{2}.$$

The principal logarithm satisfies some of the properties of the real logarithm, but we need to be careful with the addition properties.

Example 4.37. For instance, observe that

$$\operatorname{Log}(-1\cdot i)=\operatorname{Log}(-i)=\log|-i|+i\operatorname{Arg}(-i)=-i\frac{\pi}{2}.$$

On the other hand,

$$\operatorname{Log}(-1)+\operatorname{Log}i=i\pi+i\frac{\pi}{2}=i\frac{3\pi}{2},$$

so $\operatorname{Log}(-1\cdot i)\neq\operatorname{Log}(-1)+\operatorname{Log}i$. Another example of this phenomenon is

$$\operatorname{Log}(-i\cdot -i)=\operatorname{Log}(-1)=i\pi,\qquad \operatorname{Log}(-i)+\operatorname{Log}(-i)=-i\frac{\pi}{2}-i\frac{\pi}{2}=-i\pi.$$

However, under certain conditions, Log does satisfy the usual properties.

THEOREM 4.50. *Let z and w be complex numbers.*

(1) If $-\pi<\operatorname{Arg}z+\operatorname{Arg}w\leq\pi$, then

$$\operatorname{Log}zw=\operatorname{Log}z+\operatorname{Log}w.$$

(2) If $-\pi<\operatorname{Arg}z-\operatorname{Arg}w\leq\pi$, then

$$\operatorname{Log}\frac{z}{w}=\operatorname{Log}z-\operatorname{Log}w.$$

(3) If $\operatorname{Re}z,\operatorname{Re}w\geq 0$ with at least one strictly positive, then both (1) and (2) hold.

PROOF. Suppose that $-\pi<\operatorname{Arg}z+\operatorname{Arg}w\leq\pi$. By definition,

$$\operatorname{Log}zw=\log\big(|z|\,|w|\big)+i\big(\operatorname{Arg}zw\big)=\log|z|+\log|w|+i\operatorname{Arg}zw.$$

Since $\arg(zw)=\arg z+\arg w$, $\operatorname{Arg}z+\operatorname{Arg}w$ is an argument of zw, and since $\operatorname{Arg}(zw)$ is the unique argument of zw in $(-\pi,\pi]$ and $-\pi<\operatorname{Arg}z+\operatorname{Arg}w\leq\pi$, it follows that $\operatorname{Arg}(zw)=\operatorname{Arg}z+\operatorname{Arg}w$. Thus,

$$\operatorname{Log}zw=\log|z|+\log|w|+i\operatorname{Arg}z+i\operatorname{Arg}w=\operatorname{Log}z+\operatorname{Log}w.$$

Property *(2)* is prove in a similar manner. Property *(3)* follows from *(1)* and *(2)* since in case $\operatorname{Re}z,\operatorname{Re}w\geq 0$ with at least one strictly positive, then as the reader can verify, the hypotheses of both *(1)* and *(2)* are satisfied. □

We now use Log to define complex powers of *complex* numbers. Recall that given any positive real number a and complex number z, we have $a^z:=e^{z\log a}$. Using Log instead of log, we can now define powers for *complex* a. Let a be any nonzero complex number and let z be any complex number. Any number of the form e^{zb} where b is a complex logarithm of a is called a **complex power of** a **to the** z; the choice of principal logarithm defines

$$\boxed{a^z:=e^{z\operatorname{Log}a}}$$

and we call this the **principal value** of a to the power z. As before, a is called the **base** and z is called the **exponent**. Note that if a is a positive real number, then $\operatorname{Log}a=\log a$, so

$$a^z=e^{z\operatorname{Log}a}=e^{z\log a}$$

is the usual complex power of a defined in Section 4.6.3. Theorem 4.49 implies the following.

THEOREM 4.51. *The complex powers of any given nonzero complex number a to the power z are all of the form*

$$e^{z\left(\operatorname{Log}a+2\pi i\,n\right)},\qquad n\in\mathbb{Z}. \tag{4.46}$$

In general, there are infinitely many complex powers, but in certain cases they actually reduce to a finite number, see Problem 3. Here are some examples.

Example 4.38. Have you ever thought about what i^i equals? In this case, $\operatorname{Log} i = i\pi/2$, so

$$i^i = e^{i \operatorname{Log} i} = e^{i(i\pi/2)} = e^{-\pi/2},$$

a real number! Here is another nice example:

$$(-1)^{1/2} = e^{(1/2)\operatorname{Log}(-1)} = e^{(1/2)i\,\pi} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i,$$

therefore $(-1)^{1/2} = i$, just as we suspected!

4.9.4. The complex-valued arctangent function. We now investigate the complex arctangent function; the other complex inverse functions are found in Problem 5. Given a complex number z, in the following theorem we shall find all complex numbers ξ such that

$$\tan\xi = z. \tag{4.47}$$

Of course, if we can find such a ξ, then we would like to call ξ the "inverse tangent of z." However, when this equation does have solutions, it turns out that it has infinitely many.

Lemma 4.52. *If $z = \pm i$, then the equation* (4.47) *has no solutions. If $z \neq \pm i$, then*

$$\tan\xi = z \quad \Longleftrightarrow \quad e^{2i\xi} = \frac{1+iz}{1-iz},$$

that is, if and only if

$$\xi \;=\; \frac{1}{2i} \times \textit{a complex logarithm of } \frac{1+iz}{1-iz}.$$

Proof. The following statements are equivalent:

$$\tan\xi = z \quad \Longleftrightarrow \quad \sin\xi = z\cos\xi \quad \Longleftrightarrow \quad e^{i\xi} - e^{-i\xi} = iz(e^{i\xi} + e^{-i\xi})$$
$$\Longleftrightarrow \quad (e^{2i\xi} - 1) = iz(e^{2i\xi} + 1) \quad \Longleftrightarrow \quad (1 - iz)e^{2i\xi} = 1 + iz.$$

If $z = i$, then this last equation is just $2e^{2i\xi} = 0$, which is impossible, and if $z = -i$, then the last equation is $0 = 2$, again an impossibility. If $z \neq \pm i$, then the last equation is equivalent to

$$e^{2i\xi} = \frac{1+iz}{1-iz},$$

which by definition of complex logarithm just means that $2i\xi$ is a complex logarithm of the number $(1 + iz)/(1 - iz)$. □

We now choose one of the solutions of (4.47), the obvious choice being the one corresponding to the principal logarithm: Given any $z \in \mathbb{C}$ with $z \neq \pm i$, we define the **principal inverse, or arc, tangent** of z to be the complex number

$$\boxed{\operatorname{Arctan} z = \frac{1}{2i}\operatorname{Log}\frac{1+iz}{1-iz}.}$$

This defines a function $\operatorname{Arctan} : \mathbb{C} \setminus \{\pm i\} \longrightarrow \mathbb{C}$, which does satisfy (4.47):

$$\tan(\operatorname{Arctan} z) = z, \qquad z \in \mathbb{C}, \quad z \neq \pm i.$$

Some questions you might ask are whether or not Arctan really is an "inverse" of tan, in other words, is Arctan a bijection; you might also ask if $\operatorname{Arctan} x = \arctan x$ for x real. The answer to the first question is "yes," if we restrict the domain of Arctan, and the answer to the second question is "yes."

THEOREM 4.53 (**Properties of** Arctan). *Let*

$$D = \{z \in \mathbb{C}\,;\, z \neq iy,\ y \in \mathbb{R},\ |y| \geq 1\}, \qquad E = \{\xi \in \mathbb{C}\,;\, |\operatorname{Re}\xi| < \pi/2\}.$$

Then

$$\operatorname{Arctan} : D \longrightarrow E$$

is a continuous bijection from D onto E with inverse $\tan : E \longrightarrow D$ *and when restricted to real values,*

$$\operatorname{Arctan} : \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$$

and equals the usual arctangent function $\arctan : \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$.

PROOF. We begin by showing that $\operatorname{Arctan}(D) \subseteq E$. First of all, by definition of Log, for any $z \in \mathbb{C}$ with $z \neq \pm i$ (not necessarily in D) we have

$$\begin{aligned} \operatorname{Arctan} z = \frac{1}{2i} \operatorname{Log} \frac{1+iz}{1-iz} &= \frac{1}{2i}\left(\log\left|\frac{1+iz}{1-iz}\right| + i \operatorname{Arg}\frac{1+iz}{1-iz}\right) \\ &= \frac{1}{2}\operatorname{Arg}\frac{1+iz}{1-iz} - \frac{i}{2}\log\left|\frac{1+iz}{1-iz}\right|. \end{aligned} \tag{4.48}$$

Since the principal argument of any complex number lies in $(-\pi, \pi]$, it follows that

$$-\frac{\pi}{2} < \operatorname{Re}\operatorname{Arctan} z \leq \frac{\pi}{2}, \quad \text{for all } z \in \mathbb{C},\ z \neq \pm i.$$

Assume that $\operatorname{Arctan} z \notin E$, which, by the above inequality, is equivalent to

$$2\operatorname{Re}\operatorname{Arctan} z = \operatorname{Arg}\frac{1+iz}{1-iz} = \pi \quad \Longleftrightarrow \quad \frac{1+iz}{1-iz} \in (-\infty, 0).$$

If $z = x + iy$, then (by multiplying top and bottom of $\frac{1+iz}{1-iz}$ by $1 + i\overline{z}$ and making a short computation) we can write

$$\frac{1+iz}{1-iz} = \frac{1-|z|^2}{|1-iz|^2} + \frac{2\,x}{|1-iz|^2}\,i.$$

This formula shows that $(1+iz)/(1-iz) \in (-\infty, 0)$ if and only if $x = 0$ and $1 - |z|^2 < 0$, that is, $x = 0$ and $1 - y^2 < 0$, or, $|y| > 1$; hence,

$$\frac{1+iz}{1-iz} \in (-\infty, 0) \quad \Longleftrightarrow \quad z = iy\ ,\ |y| > 1.$$

In summary, for any $z \in \mathbb{C}$ with $z \neq \pm i$, we have $\operatorname{Arctan} z \notin E \Longleftrightarrow z \notin D$, or

$$\operatorname{Arctan} z \in E \quad \Longleftrightarrow \quad z \in D. \tag{4.49}$$

Therefore, $\operatorname{Arctan}(D) \subseteq E$.

We now show that $\operatorname{Arctan}(D) = E$, so let $\xi \in E$. Define $z = \tan\xi$. Then according to Lemma 4.52, we have $z \neq \pm i$ and $e^{2i\xi} = \frac{1+iz}{1-iz}$. By definition of E, the real part of ξ satisfies $-\pi/2 < \operatorname{Re}\xi < \pi/2$. Since $\operatorname{Im}(2i\xi) = 2\operatorname{Re}(\xi)$, we have $-\pi < \operatorname{Im}(2i\xi) < \pi$, and therefore by definition of the principal logarithm,

$$2i\xi = \operatorname{Log}\frac{1+iz}{1-iz}.$$

Hence, by definition of the arctangent, $\xi = \operatorname{Arctan} z$. The complex number z must be in D by (4.49) and the fact that $\xi = \operatorname{Arctan} z \in E$. This shows that $\operatorname{Arctan}(\tan \xi) = \xi$ for all $\xi \in E$ and we already know that $\tan(\operatorname{Arctan} z) = z$ for all $z \in D$ (in fact, for all $z \in \mathbb{C}$ with $z \neq \pm i$), so Arctan is a continuous bijection from D onto E with inverse given by tan.

Finally, it remains to show that Arctan equals arctan when restricted to the real line. To prove this we just need to prove that $\operatorname{Arctan} x$ is real when $x \in \mathbb{R}$. This will imply that $\operatorname{Arctan} : \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$ and therefore is just arctan. Now from (4.48) we see that if $x \in \mathbb{R}$, then the imaginary part of $\operatorname{Arctan} x$ is

$$\log \left| \frac{1+ix}{1-ix} \right| = \log \left| \frac{\sqrt{1+x^2}}{\sqrt{1+(-x)^2}} \right| = \log 1 = 0.$$

Thus, $\operatorname{Arctan} x$ is real and our proof is complete. □

Setting $z = 1$, we get Giulio Carlo Fagnano dei Toschi's (1682–1766) famous formula (see [**14**] for more on Giulio Carlo, Count Fagnano, and Marquis de Toschi):

$$\boxed{\frac{\pi}{4} = \frac{1}{2i} \operatorname{Log} \frac{1+i}{1-i}.}$$

EXERCISES 4.9.

1. Find the following logs:
$$\operatorname{Log}(1 + i\sqrt{3}), \quad \operatorname{Log}(\sqrt{3} - i), \quad \operatorname{Log}(1-i)^4,$$
and find the following powers:
$$2^i, \quad (1+i)^i, \quad e^{\operatorname{Log}(3+2i)}, \quad i^{\sqrt{3}}, \quad (-1)^{2i}.$$
2. Using trig identities, prove the following identities:
$$\arctan x + \arctan y = \arctan\left(\frac{x+y}{1-xy}\right),$$
$$\arcsin x + \arcsin y = \arcsin\left(x\sqrt{1-y^2} + y\sqrt{1-y^2}\right),$$
and give restrictions under which these identities are valid. For example, the first identity holds when $xy \neq 1$ and the left-hand side lies strictly between $-\pi/2$ and $\pi/2$. When does the second hold?
3. In this problem we study real powers of complex numbers. Let $a \in \mathbb{C}$ be nonzero.
   (a) Let $n \in \mathbb{N}$ and show that all powers of a to $1/n$ are given by
   $$e^{\frac{1}{n}\left(\operatorname{Log} a + 2\pi i\, k\right)}, \qquad k = 0, 1, 2, \ldots, n-1.$$
   In addition, show that these values are all the n-th roots of a and that the principal n-th root of a is the same as the principal value of $a^{1/n}$.
   (b) If m/n is a rational number in lowest terms with $n > 0$, show that all powers of a to m/n are given by
   $$e^{\frac{m}{n}\left(\operatorname{Log} a + 2\pi i\, k\right)}, \qquad k = 0, 1, 2, \ldots, n-1.$$
   (c) If x is an irrational number, show that there are infinitely many distinct complex powers of a to the x.
4. Let $a, b, z, w \in \mathbb{C}$ with $a, b \neq 0$ and prove the following:
   (a) $1/a^z = a^{-z}$, $a^z \cdot a^w = a^{z+w}$, and $(a^z)^n = a^{zn}$ for all $n \in \mathbb{Z}$.
   (b) If $-\pi < \operatorname{Arg} a + \operatorname{Arg} b \leq \pi$, then $(ab)^z = a^z\, b^z$.
   (c) If $-\pi < \operatorname{Arg} a - \operatorname{Arg} b \leq \pi$, then $(a/b)^z = a^z/b^z$.
   (d) If $\operatorname{Re} a, \operatorname{Re} b > 0$, then both (b) and (c) hold.

(e) Give examples showing that the conclusions of (b) and (c) are false if the hypotheses are not satisfied.

5. (**Arcsine and Arccosine function**) In this problem we define the principal arcsin and arccos functions. To define the complex arcsine, given $z \in \mathbb{C}$ we want to solve the equation $\sin \xi = z$ for ξ and call ξ the "inverse sine of z".

(a) Prove that $\sin \xi = z$ if and only if $(e^{i\xi})^2 - 2iz\,(e^{i\xi}) - 1 = 0$.

(b) Solving this quadratic equation for $e^{i\xi}$ (see Problem 5 in Exercises 4.8) prove that $\sin \xi = z$ if and only if

$$\xi \;=\; \frac{1}{i} \times \text{a complex logarithm of } iz \pm \sqrt{1 - z^2}.$$

Because of this formula, we define the **principal inverse or arc sine** of z to be the complex number

$$\boxed{\operatorname{Arcsin} z := \frac{1}{i}\,\operatorname{Log}\left(iz + \sqrt{1 - z^2}\right).}$$

Based on the formula (4.41), we define the **principal inverse or arc cosine** of z to be the complex number

$$\boxed{\operatorname{Arccos} z := \frac{\pi}{2} - \operatorname{Arcsin} z.}$$

(c) Prove that when restricted to real values, $\operatorname{Arcsin} : [-1, 1] \longrightarrow [-\pi/2, \pi/2]$ and equals the usual arcsine function.

(d) Similarly, prove that when restricted to real values, $\operatorname{Arccos} : [-1, 1] \longrightarrow [0, \pi]$ and equals the usual arccosine function.

6. (**Inverse hyperbolic functions**) We look at the inverse hyperbolic functions.

(a) Prove that $\sinh : \mathbb{R} \longrightarrow \mathbb{R}$ is a strictly increasing bijection. Thus, $\sinh^{-1} : \mathbb{R} \longrightarrow \mathbb{R}$ exists. Show that $\cosh : [0, \infty) \longrightarrow [1, \infty)$ is a strictly increasing bijection. We define $\cosh^{-1} : [1, \infty) \longrightarrow [0, \infty)$ to be the inverse of this function.

(b) Using a similar argument as you did for the arcsine function in Problem 5, prove that $\sinh x = y$ (here, $x, y \in \mathbb{R}$) if and only if $e^{2x} - 2ye^x - 1 = 0$, which holds if and only if $e^x = y \pm \sqrt{y^2 + 1}$. From this, prove that

$$\sinh^{-1} x = \log(x + \sqrt{x^2 + 1}).$$

If x is replaced by $z \in \mathbb{C}$ and log by Log, the principal complex logarithm, then this formula is called the **principal inverse hyperbolic sine** of z.

(c) Prove that

$$\cosh^{-1} x = \log(x + \sqrt{x^2 - 1}).$$

If x is replaced by $z \in \mathbb{C}$ and log by Log, the principal complex logarithm, then this formula is called the **principal inverse hyperbolic cosine** of z.

4.10. ★ The amazing π and its computations from ancient times

In the *Measurement of the circle*, Archimedes of Syracuse (287–212 B.C.), listed three famous propositions involving π (see Heath's translation [**99**]). In this section we look at each of these propositions especially his third one, which uses the first known algorithm to compute π to any desired number of decimal places![12] His basic idea is to approximate a circle by inscribed and circumscribed regular polygons. We begin by looking at a brief history of π.

[12] *[On π] Ten decimal places of are sufficient to give the circumference of the earth to a fraction of an inch, and thirty decimal places would give the circumference of the visible universe to a quantity imperceptible to the most powerful microscope. Simon Newcomb (1835–1909)* [**140**].

4.10.1. A brief (and very incomplete) history of π. We begin by giving a short snippet of the history of π with, unfortunately, many details left out. Some of what we choose to put here is based on what will come up later in this book (for example, in the chapter on continued fractions — see Section 8.5) or what might be interesting trivia. References include Schepler's comprehensive chronicles [**198, 199, 200**], the beautiful books [**26, 10, 68, 183**], the wonderful websites [**168, 170, 207**], Rice's short synopsis [**187**], and (my favorite π papers by) Castellanos [**47, 48**]. Before discussing this history, recall the following formulas for the area and circumference of a circle in terms of the radius r:

$$\text{Area of } \odot \text{ of radius } r = \pi r^2, \quad \text{Circumference of } \odot \text{ of radius } r = 2\pi r.$$

In terms of the diameter $d := 2r$, we have

$$\text{Area of } \odot = \pi \frac{d^2}{4}, \quad \text{Circumference of } \odot = \pi d, \quad \pi = \frac{\text{circumference}}{\text{diameter}}.$$

(1) (circa 1650 B.C.) The Rhind (or Ahmes) papyrus is the oldest known mathematical text in existence. It is named after the Egyptologist Alexander Henry Rhind (1833–1863) who purchased it in Luxor in 1858, but it was written by a scribe Ahmes (1680 B.C.–1620 B.C.). In this text is the following rule to find the area of a circle: *Cut $\frac{1}{9}$ off the circle's diameter and construct a square on the remainder.* Thus,

$$\pi \frac{d^2}{4} = \text{area of circle} \approx \text{square of } \Big(d - \frac{1}{9}d\Big) = \Big(d - \frac{1}{9}d\Big)^2 = \Big(\frac{8}{9}\Big)^2 d^2.$$

Cancelling off d^2 from both extremities, we obtain

$$\pi \approx 4\Big(\frac{8}{9}\Big)^2 = \Big(\frac{4}{3}\Big)^4 = 3.160493827\ldots.$$

(2) (circa 1000 B.C.) The Holy Bible in I Kings, Chapter 7, verse 23, and II Chronicles, Chapter 4, verse 2, states that

> And he made a molten sea, ten cubits from the one brim to the other:
> it was round all about, and his height was five cubits: and a line of
> thirty cubits did compass it about. I Kings 7:23.

This give the approximate value (cf. the interesting article [**5**]):

$$\pi = \frac{\text{circumference}}{\text{diameter}} \approx \frac{30 \text{ cubits}}{10 \text{ cubits}} = 3.$$

Not only did the Israelites use 3, other ancient civilizations used 3 for rough purposes (more than good enough for "everyday life") like the Babylonians, Hindus, and Chinese.

(3) (circa 250 B.C.) Archimedes of Syracuse (287–212) gave the estimate $\pi \approx 22/7 = 3.14285714\ldots$ (correct to two decimal places). We'll thoroughly discuss "Archimedes' method" in a moment.

(4) (circa 500 A.D.) Tsu Chung-Chi (also Zu Chongzhi) of China (429–501) gave the estimate $\pi \approx 355/113 = 3.14159292\ldots$ (correct to six decimal places); he also gave the incredible estimate

$$3.1415926 < \pi < 3.1415927.$$

(5) (circa 1600 A.D.) The Dutch mathematician Adriaan Anthoniszoon of Holland (1527–1607) used Archemides' method to get

$$\frac{333}{106} < \pi < \frac{377}{120}.$$

By taking the average of the numerators and the denominator, he found Tsu Chung-Chi's approximation 355/113.

(6) (1706) The symbol π was first introduced by William Jones (1675–1749) in his beginners calculus book *Synopsis palmariorum mathesios* where he published John Machin's (1680–1751) one hundred digit approximation to π; see Subsection 4.10.5 for more on Machin. The symbol π was popularized and became standard through Leonhard Euler's (1707–1783) famous book *Introductio in Analysin Infinitorum* [**65**]. The letter π was (reportedly) chosen because it's the first letter of the Greek words "perimeter" and "periphery".

(7) (1761) Johann Heinrich Lambert (1728–1777) proved that π is irrational.

(8) (1882) Carl Louis Ferdinand von Lindemann (1852–1939) proved that π is transcendental.

(9) (1897) A sad day in the life of π. House bill No. 246, Indiana state legislature, 1897, written by a physician Edwin Goodwin (1828–1902), tried to legally set the value of π to a rational number; see [**213**], [**90**] for more about this sad tale. This value would be copyrighted and used in Indiana state math textbooks and other states would have to pay to use this value! The bill is very convoluted (try to read Goodwin's article [**83**] and you'll probably get a headache) and (reportedly) the following values of π can be implied from the bill: $\pi = 9.24$, 3.236, 3.232, and 3.2; it's also implied that $\sqrt{2} = 10/7$. Moreover, Mr. Goodwin claimed he could trisect an angle, double a cube, and square a circle, which (quoting from the bill) "had been long since given up by scientific bodies as insolvable mysteries and above mans ability to comprehend." These problems "had been long since given up" because they have been proven unsolvable! (See [**57, 79**] for more on these unsolvable problems, first proved by Pierre Laurent Wantzel (1814–1848), and see [**59**] for other stories of amateur mathematicians claiming to have solved the insolvable.) This bill passed the house (!), but fortunately, with the help of mathematician C.A. Waldo of the Indiana Academy of Science, the bill didn't pass in the senate.

Hold on to your seats because we'll take up our brief history of π again in Subsection 4.10.5, after a brief intermission.

4.10.2. Archimedes' three propositions.

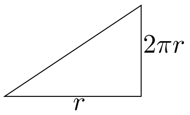
The following three propositions are contained in Archimedes' book *Measurement of the circle* [**99**]:

(1) The area of a circle is equal to that of a right-angled triangle where the sides including the right angle are respectively equal to the radius and circumference of the circle.

(2) The ratio of the area of a circle to that of a square with side equal to the circle's diameter is close to 11:14.

(3) The ratio of the circumference of any circle to its diameter is less than 3 1/7 but greater than 3 10/71.

Archimedes' first proposition is seen in Figure 4.13. Archimedes' second propo-



$$\text{area } \triangle = \frac{1}{2}\text{base} \times \text{height} = \frac{1}{2} r \cdot (2\pi r) = \pi r^2$$

FIGURE 4.13. Archimedes' first proposition.

sition gives the famous estimate $\pi \approx \frac{22}{7}$:

$$\frac{\text{area of circle}}{\text{area of square}} = \frac{\pi r^2}{(2r)^2} = \frac{\pi}{4} \approx \frac{11}{14} \quad \Longrightarrow \quad \pi \approx \frac{22}{7}.$$

We now derive Archimedes' third proposition using the same method Archimedes pioneered over two thousand years ago, but we shall employ trigonometric functions! Archimedes' original method used plane geometry to derive his formulas (they didn't have the knowledge of trigonometric functions back then as we do now.) However, before doing so, we need a couple trig facts.

4.10.3. Some useful trig facts.

We first consider some useful trig identities.

LEMMA 4.54. *We have*

$$\tan z = \frac{\sin(2z)\tan(2z)}{\sin(2z) + \tan(2z)} \quad \textit{and} \quad 2\sin^2 z = \sin(2z)\tan z.$$

PROOF. We'll prove the first one and leave the second one to you. Multiplying $\tan z$ by $2\cos z/2\cos z = 1$ and using the double angle formulas $2\cos^2 z = 1 + \cos 2z$ and $\sin(2z) = 2\cos z \sin z$ (see Theorem 4.34), we obtain

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

Multiplying top and bottom by $\tan 2z$, we get

$$\tan z = \frac{\sin(2z)\tan(2z)}{\tan(2z) + \cos(2z)\tan 2z} = \frac{\sin(2z)\tan(2z)}{\tan(2z) + \sin(2z)}.$$

□

Next, we consider some useful inequalities.

LEMMA 4.55. *For $0 < x < \pi/2$, we have*

$$\sin x < x < \tan x.$$

PROOF. We first prove that $\sin x < x$ for $0 < x < \pi/2$. We note that the inequality $\sin x < x$ for $0 < x < \pi/2$ automatically implies that this same inequality holds for all $x > 0$, since x is increasing and $\sin x$ is oscillating. Substituting the power series for $\sin x$, the inequality $\sin x < x$, that is, $-x < -\sin x$, is equivalent to

$$-x < -x + \frac{x^3}{3!} - \frac{x^5}{5!} + \frac{x^7}{7!} - \frac{x^9}{9!} + - \cdots,$$

or after cancelling off the x's, this inequality is equivalent to

$$\frac{x^3}{3!}\left(1 - \frac{x^2}{4 \cdot 5}\right) + \frac{x^7}{7!}\left(1 - \frac{x^2}{8 \cdot 9}\right) + \cdots > 0.$$

For $0 < x < 2$, each of the terms in parentheses is positive. This shows that in particular, this expression is positive for $0 < x < \pi/2$.

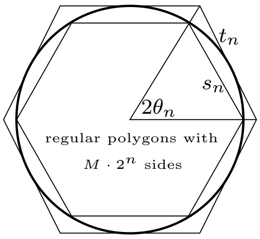


FIGURE 4.14. Archimedes inscribed and circumscribed a circle with diameter 1 (radius 1/2) with regular polygons. The sides of these polygons have lengths s_n and t_n, respectively. $2\theta_n$ is the central angle of the inscribed and circumscribed $2^n M$-gons.

We now prove that $x < \tan x$ for $0 < x < \pi/2$. This inequality is equivalent to $x \cos x < \sin x$ for $0 < x < \pi/2$. Substituting the power series for cos and sin, the inequality $x \cos x < \sin x$ is equivalent to

$$x - \frac{x^3}{2!} + \frac{x^5}{4!} - \frac{x^7}{6!} + - \cdots < x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + - \cdots .$$

Bringing everything to the right, we get an inequality of the form

$$x^3\left(\frac{1}{2!} - \frac{1}{3!}\right) - x^5\left(\frac{1}{4!} - \frac{1}{5!}\right) + x^7\left(\frac{1}{6!} + \frac{1}{7!}\right) - x^9\left(\frac{1}{8!} + \frac{1}{9!}\right) + - \cdots > 0.$$

Combining adjacent terms, the left-hand side is a sum of terms of the form

$$x^{2k-1}\left(\frac{1}{(2k-2)!} - \frac{1}{(2k-1)!}\right) - x^{2k+1}\left(\frac{1}{(2k)!} - \frac{1}{(2k+1)!}\right), \quad k = 2, 3, 4, \cdots .$$

We claim that this term is positive for $0 < x < 3$. This shows that $x \cos x < \sin x$ for $0 < x < 3$, and so in particular, for $0 < x < \pi/2$. Now the above expression is positive if and only if

$$x^2 < \frac{\dfrac{1}{(2k-2)!} - \dfrac{1}{(2k-1)!}}{\dfrac{1}{(2k)!} - \dfrac{1}{(2k+1)!}} = (2k+1)(2k-2), \quad k = 2, 3, 4, \ldots .$$

where we multiplied the top and bottom by $(2k+1)!$. The right-hand side is smallest when $k = 2$, when it equals $5 \cdot 2 = 10$. It follows that these inequalities hold for $0 < x < 3$, and our proof is now complete. □

4.10.4. Archimedes' third proposition. We start with a circle with diameter 1 (radius 1/2). Then,

$$\text{circumference of this circle} = 2\pi r = 2\pi\Big(\frac{1}{2}\Big) = \pi.$$

Let us fix a natural number $M \geq 3$. Given any $n = 0, 1, 2, 3, \ldots$, we inscribe and circumscribe the circle with regular polygons having $2^n M$ sides. See Figure 4.14. We denote the perimeter of the inscribed $2^n M$-gon by lowercase p_n and the perimeter of the circumscribed $2^n M$-gon by uppercase P_n. Then geometrically we can see that

$$p_n < \pi < P_n, \qquad n = 0, 1, 2, \ldots$$

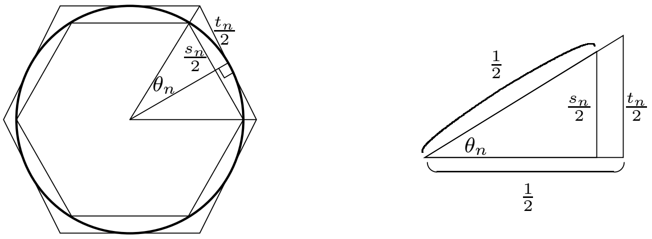


FIGURE 4.15. We cut the central angle in half. The right picture shows a blow-up of the overlapping triangles on the left.

and $p_n \to \pi$ and $P_n \to \pi$ as $n \to \infty$; we shall prove these facts analytically in Theorem 4.56. Using plane geometry, Archimedes found iterative formulas for the sequences $\{p_n\}$ and $\{P_n\}$ and using these formulas he proved his third proposition. Recall that everything we thought about trigonometry is true ☺, so we shall use these trig facts to derive Archimedes' famous iterative formulas for the sequences $\{p_n\}$ and $\{P_n\}$. To this end, we let s_n and t_n denote the length of the sides of the inscribed and circumscribed polygons, so that

$$P_n = (\#\text{ sides}) \times (\text{length each side}) = 2^n M \cdot t_n.$$

and

$$p_n = (\#\text{ sides}) \times (\text{length each side}) = 2^n M \cdot s_n$$

Let $2\theta_n$ be the central angle of the inscribed and circumscribed $2^n M$-gons as shown in Figures 4.14 and 4.15 (that is, θ_n is half the central angle). The right picture in Figure 4.15 gives a blown-up picture of the triangles in the left-hand picture. The outer triangle in the middle picture shows that

$$\tan\theta_n = \frac{\text{opposite}}{\text{adjacent}} = \frac{t_n/2}{1/2} \quad\Longrightarrow\quad t_n = \tan\theta_n \quad\Longrightarrow\quad P_n = 2^n M \tan\theta_n.$$

The inner triangle shows that

$$\sin\theta_n = \frac{\text{opposite}}{\text{hypotonus}} = \frac{s_n/2}{1/2} \quad\Longrightarrow\quad s_n = \sin\theta_n \quad\Longrightarrow\quad p_n = 2^n M \sin\theta_n.$$

Now what's θ_n? Well, since $2\theta_n$ is the central angle of the $2^n M$-gon, we have

$$\text{central angle} = \frac{\text{total angle of circle}}{\#\text{ of sides of regular polygon}} = \frac{2\pi}{2^n M}.$$

Setting this equal to $2\theta_n$, we get

$$\theta_n = \frac{\pi}{2^n M}.$$

In particular,

$$\theta_{n+1} = \frac{\pi}{2^{n+1} M} = \frac{1}{2}\frac{\pi}{2^n M} = \frac{1}{2}\theta_n.$$

Setting z equal to $z/2$ in Lemma 4.54, we see that

$$\tan\left(\frac{1}{2}z\right) = \frac{\sin(z)\tan(z)}{\sin(z)+\tan(z)} \quad \text{and} \quad 2\sin^2\left(\frac{1}{2}z\right) = \sin(z)\tan\left(\frac{1}{2}z\right).$$

Hence,

$$\tan\theta_{n+1} = \tan\left(\frac{1}{2}\theta_n\right) = \frac{\sin(\theta_n)\tan(\theta_n)}{\sin(\theta_n)+\tan(\theta_n)}$$

and

$$2\sin^2(\theta_{n+1}) = 2\sin^2\left(\frac{1}{2}\theta_n\right) = \sin(\theta_n)\tan\left(\frac{1}{2}\theta_n\right) = \sin(\theta_n)\tan(\theta_{n+1}).$$

In particular, recalling that $P_n = 2^n M \tan\theta_n$ and $p_n = 2^n M \sin\theta_n$, we see that

$$\begin{aligned} P_{n+1} = 2^{n+1}M\tan\theta_{n+1} &= 2^{n+1}M\frac{\sin(\theta_n)\tan(\theta_n)}{\sin(\theta_n)+\tan(\theta_n)} \\ &= 2\frac{2^nM\sin(\theta_n)\cdot 2^nM\tan(\theta_n)}{2^nM\sin(\theta_n)+2^nM\tan(\theta_n)} = 2\frac{p_nP_{n+1}}{p_n+P_n}. \end{aligned}$$

and

$$\begin{aligned} 2p_{n+1}^2 = 2\big(2^{n+1}M\sin\theta_{n+1}\big)^2 &= \big(2^{n+1}M\big)^2\sin(\theta_n)\tan(\theta_{n+1}) \\ &= 2\cdot 2^nM\sin(\theta_n)\,2^{n+1}M\tan(\theta_{n+1}) = 2p_nP_{n+1}, \end{aligned}$$

or $p_{n+1} = \sqrt{p_n\,P_{n+1}}$. Finally, recall that $\theta_n = \frac{\pi}{2^nM}$. Thus, $P_0 = M\tan(\frac{\pi}{M})$ and $p_0 = M\sin(\frac{\pi}{M})$. Let us summarize our results in the following formulas:

$$(4.50)\qquad \boxed{\begin{array}{l} P_{n+1} = \dfrac{2p_nP_n}{p_n+P_n}, \quad p_{n+1} = \sqrt{p_n\,P_{n+1}}; \qquad \textbf{(Archimedes' algorithm)} \\ P_0 = M\tan\left(\dfrac{\pi}{M}\right), \quad p_0 = M\sin\left(\dfrac{\pi}{M}\right). \end{array}}$$

This is the celebrated Archimedes' algorithm. Starting from the values of P_0 and p_0, we can use the iterative definitions for P_{n+1} and p_{n+1} to generate sequences $\{P_n\}$ and $\{p_n\}$ that converge to π, as we now show.

THEOREM 4.56 (**Archimedes' algorithm**). *We have*

$$p_n < \pi < P_n, \qquad n = 0, 1, 2, \ldots$$

and $p_n \to \pi$ and $P_n \to \pi$ as $n \to \infty$.

PROOF. Note that for any $n = 0, 1, 2, \ldots$, we have $0 < \theta_n = \frac{\pi}{2^nM} < \frac{\pi}{2}$ because $M \geq 3$. Thus, by Lemma 4.55,

$$p_n = 2^nM\sin\theta_n < 2^nM\theta_n < 2^nM\tan\theta_n = P_n.$$

Since $\theta_n = \frac{\pi}{2^nM}$, the middle term is just π, so $p_n < \pi < P_n$ for every $n = 0, 1, 2, \ldots$. Using the limit $\lim_{z\to 0}\sin z/z = 1$, we obtain

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} 2^nM\sin\theta_n = \lim_{n\to\infty} \pi\frac{\sin\left(\frac{\pi}{2^nM}\right)}{\left(\frac{\pi}{2^nM}\right)} = \pi.$$

Since $\lim_{z\to 0}\cos z = 1$, we have $\lim_{z\to 0}\tan z/z = \lim_{z\to 0}\sin z/(z\cdot\cos z) = 1$, so the same argument we used for p_n shows that $\lim_{n\to\infty} P_n = \pi$. □

In Problem 4 you will study how fast p_n and P_n converge to π. Now let's consider a specific example: Let $M = 6$, which is what Archimedes chose! Then,

$$P_0 = 6\tan\left(\frac{\pi}{6}\right) = 2\sqrt{3} = 3.464101615\ldots \quad \text{and} \quad p_0 = 6\sin\left(\frac{\pi}{6}\right) = 3.$$

From these values, we can find P_1 and p_1 from Archimedes algorithm (4.50):

$$P_1 = \frac{2p_0P_0}{p_0+P_0} = \frac{2\cdot 3\cdot 2\sqrt{3}}{3+2\sqrt{3}} = 3.159659942\ldots$$

and

$$p_1 = \sqrt{p_0\, P_1} = \sqrt{3 \cdot 3.159659942\ldots} = 3.105828541\ldots.$$

Continuing this process (I used a spreadsheet) we can find P_2, p_2, then P_3, p_3, and so forth, arriving with the table

n	p_n	P_n
0	3	3.464101615
1	3.105828541	3.215390309
2	3.132628613	3.159659942
3	3.139350203	3.146086215
4	3.141031951	3.1427146
5	3.141452472	3.14187305
6	3.141557608	3.141662747
7	3.141583892	3.141610177

Archimedes considered $p_4 = 3.14103195\ldots$ and $P_4 = 3.1427146\ldots$. Notice that

$$3\frac{10}{71} = 3.140845070\ldots < p_4 \quad \text{and} \quad P_4 < 3.142857142\ldots = 3\frac{1}{7}.$$

Hence,

$$3\frac{10}{71} < p_4 < \pi < P_4 < 3\frac{1}{7},$$

which proves Archimedes' third proposition. It's interesting to note that Archimedes didn't have computers back then (to find square roots for instance), or trig functions, or coordinate geometry, or decimal notation, etc. so it's incredible that Archimedes was able to determine π to such an incredible accuracy!

4.10.5. Continuation of our brief history of π. Here are (only some!) famous formulas for π (along with their earliest known date of publication) that we'll prove in our journey through our book: Archimedes of Syracuse ≈ 250 B.C.: $\pi = \lim P_n = \lim p_n$, where

$$\boxed{P_{n+1} = \frac{2p_nP_n}{p_n + P_n}, \quad p_{n+1} = \sqrt{p_n\, P_{n+1}}; \quad P_0 = M \tan\Big(\frac{\pi}{M}\Big), \quad p_0 = M \sin\Big(\frac{\pi}{M}\Big).}$$

We remark that Archimedes' algorithm is similar to Borchardt's algorithm (see Problem 1), which is similar to the modern-day AGM method of Eugene Salamin, Richard Brent, and Jonathan and Peter Borwein [**32, 33**]. This AGM method can generate *billions* of digits of π!

François Viète 1593 (§ 5.1):

$$\boxed{\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots.}$$

Lord William Brouncker 1655 (§ 7.7):

$$\boxed{\frac{4}{\pi} = 1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \cdots}}}}.}$$

John Wallis 1656 (§ 6.10):

$$\boxed{\frac{\pi}{2} = \prod_{n=1}^{\infty} \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots .}$$

James Gregory, Gottfried Wilhelm von Leibniz 1670, Madhava of Sangamagramma $\approx$ 1400 (§ 5.1):

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + - \cdots .}$$

John Machin 1706 (§ 6.10):

$$\boxed{\pi = 4\arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right) = 4\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)} \left(\frac{4}{5^{2n+1}} - \frac{1}{239^{2n+1}}\right).}$$

Machin calculated 100 digits of π with this formula. William Shanks (1812–1882) is famed for his calculation of π to 707 places in 1873 using Machin's formula. However, only the first 527 places were correct as discovered by D. Ferguson in 1944 [**72**] using another Machin type formula. Ferguson ended up publishing 620 correct places in 1946, which marks the last hand calculation for π ever to so many digits. From this point on, computers have been used to find π and the number of digits of π known today is well into the *trillions* lead by Yasumasa Kanada and his coworkers at the University of Tokyo using a Machin type formula; see Kanada's website http://www.super-computing.org/. One might ask "why try to find so many digits of π?" Well (taken from Young's great book [**252**, p. 238]),

> *Perhaps in some far distant century they may say, "Strange that those ingenious investigators into the secrets of the number system had so little conception of the fundamental discoveries that would later develop from them!" D. N. Lehmer (1867–1938).*

We now go back to our list of formulas. Leonhard Euler 1736 (§ 5.1):

$$\boxed{\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots .}$$

and (§ 5.1):

$$\boxed{\frac{\pi^2}{6} = \frac{2^2}{2^2-1} \cdot \frac{3^2}{3^2-1} \cdot \frac{5^2}{5^2-1} \cdot \frac{7^2}{7^2-1} \cdot \frac{11^2}{11^2-1} \cdots .}$$

We end our history with a question to ponder: What is the probability that a natural number, chosen at random, is square free (that is, is not divisible by the square of a prime)? What is the probability that two natural numbers, chosen at random, are relatively (or co-) prime (that is, don't have any common prime factors)? The answers, drum role please, (§ 7.6):

$$\boxed{\text{Probability of being square free} = \text{Probability of being coprime} = \frac{6}{\pi^2}.}$$

Exercises 4.10.

1. In a letter from Gauss to his teacher Johann Pfaff (1765–1825) around 1800, Gauss asked Pfaff about the following sequences $\{\alpha_n\}$, $\{\beta_n\}$ defined recursively as follows:
$$\alpha_{n+1} = \frac{1}{2}(\alpha_n + \beta_n), \quad \beta_{n+1} = \sqrt{\alpha_{n+1}\beta_n}. \quad \textbf{(Borchardt's algorithm)}$$
Later, Carl Borchardt (1817–1880) rediscovered this algorithm and since then this algorithm is called **Borchardt's algorithm** [**46**]. Prove that Borchardt's algorithm is basically the same as Archimedes' algorithm in the following sense: if you set $\alpha_n := 1/P_n$ and $\beta_n := 1/p_n$ in Archimedes' algorithm, you get Borchardt's algorithm.
2. (**Pfaff's solution I**) Now what if we don't use the starting values $P_0 = M\tan\left(\frac{\pi}{M}\right)$ and $p_0 = M\sin\left(\frac{\pi}{M}\right)$ for Archimedes' algorithm in (4.50), but instead used other starting values? What do the sequences $\{P_n\}$ and $\{p_n\}$ converge to? These questions were answered by Johann Pfaff. Pick starting values P_0 and p_0 and let's assume that $0 \le p_0 < P_0$; the case that $P_0 < p_0$ is handled in the next problem.
   (i) Define
$$\theta := \arccos\left(\frac{p_0}{P_0}\right), \qquad r := \frac{p_0 P_0}{\sqrt{P_0^2 - p_0^2}}.$$
   Prove that $P_0 = r\tan\theta$ and $p_0 = r\sin\theta$.
   (ii) Prove by induction that $P_0 = 2^n r\tan\left(\frac{\theta}{2^n}\right)$ and $p_n = 2^n r\sin\left(\frac{\theta}{2^n}\right)$.
   (iii) Prove that as $n \to \infty$, both $\{P_n\}$ and $\{p_n\}$ converge to
$$r\theta = \frac{p_0 P_0}{\sqrt{P_0^2 - p_0^2}} \arccos\left(\frac{p_0}{P_0}\right).$$
3. (**Pfaff's solution II**) Now assume that $0 < P_0 < p_0$.
   (i) Define (see Problem 6 in Exercises 4.9 for the definition of $\cosh^{-1}$)
$$\theta := \cosh^{-1}\left(\frac{p_0}{P_0}\right), \qquad r := \frac{p_0 P_0}{\sqrt{p_0^2 - P_0^2}}.$$
   Prove that $P_0 = r\tanh\theta$ and $p_0 = r\sinh\theta$.
   (ii) Prove by induction that $P_0 = 2^n r\tanh\left(\frac{\theta}{2^n}\right)$ and $p_n = 2^n r\sinh\left(\frac{\theta}{2^n}\right)$.
   (iii) Prove that as $n \to \infty$, both $\{P_n\}$ and $\{p_n\}$ converge to
$$r\theta = \frac{p_0 P_0}{\sqrt{p_0^2 - P_0^2}} \cosh^{-1}\left(\frac{p_0}{P_0}\right).$$
4. (Cf. [**150**], [**181**]) (Rate of convergence)
   (a) Using the formulas $p_n = 2^n M\sin\theta_n$ and $P_n = 2^n M\tan\theta_n$, where $\theta_n = \frac{\pi}{2^n M}$, prove that there are constants $C_1, C_2 > 0$ such that for all n,
$$|p_n - \pi| \le \frac{C_1}{4^n} \quad \text{and} \quad |P_n - \pi| \le \frac{C_2}{4^n}.$$
   Suggestion: For the first estimate, use the expansion $\sin z = z - \frac{z^3}{3!} + \cdots$. For the second estimate, notice that $|P_n - \pi| = \frac{1}{\cos\theta_n}|p_n - \pi\cos\theta_n|$.
   (b) Part (a) shows that $\{p_n\}$ and $\{P_n\}$ converge to π very fast, but we can get even faster convergence by looking at the sequence $\{a_n\}$ where $a_n := \frac{1}{3}(2p_n + P_n)$. Prove that there is a constant $C > 0$ such that for all n,
$$|a_n - \pi| \le \frac{C}{16^n}.$$

CHAPTER 5

Some of the most beautiful formulæ in the world

God used beautiful mathematics in creating the world.
Paul Adrien Maurice Dirac (1902–1984)

In this chapter we present a small sample of *some* of the most beautiful formulas in the world. We begin in Section 5.1 where we present Viète's formula, Wallis' formula, and Euler's sine expansion. Viète's formula, due to François Vite (1540–1603), is the infinite product

$$\boxed{\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots ,}$$

published in 1593. This is not only the first recorded infinite product [**120**, p. 218] it is also the first recorded theoretically *exact* analytical expression for the number π [**36**, p. 321]. Wallis' formula, named after John Wallis (1616–1703) was the second recorded infinite product [**120**, p. 219]:

$$\boxed{\frac{\pi}{2} = \prod_{n=1}^{\infty} \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots .}$$

To explain Euler's sine expansion, recall that if $p(x)$ is a polynomial with nonzero roots $r_1, \ldots, r_n$ (repeated according to multiplicity), then we can factor $p(x)$ as $p(x) = a(x-r_1)(x-r_2)\cdots(x-r_n)$ where a is a constant. Factoring out $-r_1, \ldots, -r_n$, we can write $p(x)$ as

$$p(x) = b\left(1 - \frac{x}{r_1}\right)\left(1 - \frac{x}{r_2}\right) \cdots \left(1 - \frac{x}{r_n}\right), \tag{5.1}$$

for another constant b. Euler noticed that the function $\frac{\sin x}{x}$ has only nonzero roots, located at

$$\pi, -\pi, 2\pi, -2\pi, 3\pi, -3\pi, \ldots,$$

so thinking of $\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \cdots$ as a (infinite) polynomial, assuming that (5.1) holds for such an infinite polynomial we have (without caring about being rigorous for the moment!)

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

where b is a constant. In Section 5.1, we prove that Euler's guess was correct (with $b = 1$)! Here's Euler's famous formula:

$$\boxed{\sin x = x\Big(1 - \frac{x^2}{\pi^2}\Big)\Big(1 - \frac{x^2}{2^2\pi^2}\Big)\Big(1 - \frac{x^2}{3^2\pi^2}\Big)\Big(1 - \frac{x^2}{4^2\pi^2}\Big)\Big(1 - \frac{x^2}{5^2\pi^2}\Big)\cdots,} \tag{5.2}$$

which Euler proved in 1735 in his epoch-making paper *De summis serierum reciprocarum* (On the sums of series of reciprocals), which was read in the St. Petersburg Academy on December 5, 1735 and originally published in Commentarii academiae scientiarum Petropolitanae 7, 1740, and reprinted on pp. 123–134 of Opera Omnia: Series 1, Volume 14, pp. 73–86.

In Section 5.2 we study the Basel problem, which is the problem to determine the exact value of $\zeta(2) = \sum_{n=1}^\infty \frac{1}{n^2}$. Euler, in the same 1735 paper *De summis serierum reciprocarum* proved that $\zeta(2) = \frac{\pi^2}{6}$:

$$\boxed{\sum_{n=1}^\infty \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}.}$$

Euler actually gave three proofs of this formula in *De summis serierum reciprocarum*, but the third one is the easiest to explain. Here it is: First, recall Euler's sine expansion:

$$\frac{\sin x}{x} = \left(1 - \frac{x^2}{1^2\pi^2}\right)\left(1 - \frac{x^2}{2^2\pi^2}\right)\left(1 - \frac{x^2}{3^2\pi^2}\right)\left(1 - \frac{x^2}{4^2\pi^2}\right)\cdots.$$

If you think about multiply out the right-hand side you will get

$$1 - \frac{x^2}{\pi^2}\left(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots\right) + \cdots$$

where the dots "$\cdots$" involves powers of x of degree at least four or higher. Thus,

$$\frac{\sin x}{x} = 1 - \frac{x^2}{\pi^2}\zeta(2) + \cdots.$$

Dividing the power series of $\sin x = x - \frac{x^3}{3!} + \cdots$ by x we conclude that

$$1 - \frac{x^2}{3!} + \cdots = 1 - \frac{x^2}{\pi^2}\zeta(2) + \cdots$$

where "$\cdots$" involves powers of x of degree at least four or higher. Finally, equating powers of x^2 we conclude that

$$-\frac{1}{3!} = \frac{\zeta(2)}{\pi^2} \quad \Longrightarrow \quad \zeta(2) = \frac{\pi^2}{3!} = \frac{\pi^2}{6}.$$

Here is Jordan Bell's [**21**] English translation of Euler's argument from *De summis serierum reciprocarum* (which was originally written in Latin):

> *Indeed, it having been put*[1] $y = 0$, *from which the fundamental equation will turn into this*[2]
>
> $$0 = s - \frac{s^3}{1\cdot 2\cdot 3} + \frac{s^5}{1\cdot 2\cdot 3\cdot 4\cdot 5} - \frac{s^7}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7} + \mathit{etc.},$$

[1] Here, Euler set $y = \sin s$.

[2] Instead of writing e.g. $1\cdot 2\cdot 3$, today we would write this as 3!. However, the factorial symbol wasn't invented until 1808 [**151**], by Christian Kramp (1760–1826), more than 70 years after *De summis serierum reciprocarum* was read in the St. Petersburg Academy.

> *The roots of this equation give all the arcs of which the sine is equal to* 0*. Moreover, the single smallest root is $s = 0$, whereby the equation divided by s will exhibit all the remaining arcs of which the sine is equal to* 0*; these arcs will hence be the roots of this equation*
>
> $$0 = 1 - \frac{s^2}{1\cdot 2\cdot 3} + \frac{s^4}{1\cdot 2\cdot 3\cdot 4\cdot 5} - \frac{s^6}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7} + etc.$$
>
> *Truly then, those arcs of which the sine is equal to* 0 *are*[3]
>
> $$p\,, \quad -p\,, \quad +2p\,, \quad -2p\,, \quad 3p\,, \quad -3p \quad etc.,$$
>
> *of which the the second of the two of each pair is negative, each of these because the equation indicates for the dimensions of s to be even. Hence the divisors of this equation will be*
>
> $$1 - \frac{s}{p}\,, \quad 1 + \frac{s}{p}\,, \quad 1 - \frac{s}{2p}\,, \quad 1 + \frac{s}{2p} \quad etc.$$
>
> *and by the joining of these divisors two by two it will be*
>
> $$\begin{aligned} 1 - \frac{s^2}{1\cdot 2\cdot 3} &+ \frac{s^4}{1\cdot 2\cdot 3\cdot 4\cdot 5} - \frac{s^6}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7} + etc. \\ &= \left(1 - \frac{s^2}{p^2}\right)\left(1 - \frac{s^2}{4p^2}\right)\left(1 - \frac{s^2}{9p^2}\right)\left(1 - \frac{s^2}{16p^2}\right) \quad etc. \end{aligned}$$
>
> *It is now clear from the nature of equations for the coefficient*[4] *of ss that is $\frac{1}{1\cdot 2\cdot 3}$ to be equal to*
>
> $$\frac{1}{p^2} + \frac{1}{4p^2} + \frac{1}{9p^2} + \frac{1}{16p^2} + etc.$$

In this last step, Euler says that

$$\frac{1}{1\cdot 2\cdot 3} = \frac{1}{p^2} + \frac{1}{4p^2} + \frac{1}{9p^2} + \frac{1}{16p^2} + \text{etc},$$

which after some rearrangement is exactly the statement that $\zeta(2) = \frac{\pi^2}{6}$. Euler's proof reminds me of a quote by Charles Hermite (1822–1901):

> *There exists, if I am not mistaken, an entire world which is the totality of mathematical truths, to which we have access only with our mind, just as a world of physical reality exists, the one like the other independent of ourselves, both of divine creation. Quoted in The Mathematical Intelligencer, vol. 5, no. 4.*

By the way, in this book we give eleven proofs of Euler's formula for $\zeta(2)$!

In Section 5.2 we also prove the Gregory-Leibniz-Madhava series

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots .}$$

This formula is usually called **Leibniz's series** after Gottfried Leibniz (1646–1716) because he is usually accredited to be the first to mention this formula in print in 1673, although James Gregory (1638–1675) probably knew about it. However, the

[3] Here, Euler uses p for π. The notation π for the ratio of the length of a circle to its diameter was introduced in 1706 by William Jones (1675–1749), and only around 1736, a year after Euler published *De summis serierum reciprocarum*, did Euler seem to adopt the notation π.

[4] Here, ss means s^2.

great mathematician and astronomer Madhava of Sangamagramma (1350–1425) from India discovered this formula over 200 years before either Gregory or Leibniz!

Finally, in Section 5.3 we derive Euler's formula for $\zeta(n)$ for all even n.

Chapter 5 objectives: The student will be able to ...

- Explain the various formulas of Euler, Wallis, Viète, Gregory, Leibniz, Madhava.
- Formally derive Euler's sine expansion and formula for $\pi^2/6$.
- describe Euler's formulæ for $\zeta(n)$ for n even.

5.1. ★ Euler, Wallis, and Viète

Historically, Viète's formula was the first infinite product written down and Wallis' formula was the second [**120**, p. 218–219]. In this section we prove these formulas and we also prove Euler's celebrated sine expansion.

5.1.1. Viète's Formula: The first analytic expression for π.

François Viète's (1540–1603) formula has a very elementary proof. For any nonzero $z \in \mathbb{C}$, dividing the identity $\sin z = 2\sin(z/2)\cos(z/2)$ by z, we get

$$\frac{\sin z}{z} = \cos(z/2) \cdot \frac{\sin(z/2)}{z/2}.$$

Replacing z with $z/2$, we get $\sin(z/2)/(z/2) = \cos(z/2^2)\cdot\sin(z/2^2)/(z/2^2)$, therefore

$$\frac{\sin z}{z} = \cos(z/2) \cdot \cos(z/2^2) \cdot \frac{\sin(z/2^2)}{z/2^2}.$$

Continuing by induction, we obtain

$$\begin{aligned} \frac{\sin z}{z} &= \cos(z/2) \cdot \cos(z/2^2) \cdots \cos(z/2^n) \cdot \frac{\sin(z/2^n)}{z/2^n} \\ &= \frac{\sin(z/2^n)}{z/2^n} \cdot \prod_{k=1}^{n} \cos(z/2^k), \end{aligned} \tag{5.3}$$

or

$$\prod_{k=1}^{n} \cos(z/2^k) = \frac{z/2^n}{\sin(z/2^n)} \cdot \frac{\sin z}{z}.$$

Since $\lim_{n\to\infty} \frac{z/2^n}{\sin(z/2^n)} = 1$ for any nonzero $z \in \mathbb{C}$, we have

$$\lim_{n\to\infty} \prod_{k=1}^{n} \cos(z/2^k) = \lim_{n\to\infty} \frac{\sin z}{z} \cdot \frac{z/2^n}{\sin(z/2^n)} = \frac{\sin z}{z}.$$

For notation purposes, we can write

$$\frac{\sin z}{z} = \prod_{n=1}^{\infty} \cos(z/2^n) = \cos(z/2) \cdot \cos(z/2^2) \cdot \cos(z/2^4) \cdots \tag{5.4}$$

and refer the right-hand side as an **infinite product**, the subject of which we'll thoroughly study in Chapter 7. For the purposes of this chapter, given a sequence $a_1, a_2, a_3, \ldots$ we shall denote by $\prod_{n=1}^{\infty} a_n$ as the limit

$$\prod_{n=1}^{\infty} a_n := \lim_{n\to\infty} \prod_{k=1}^{n} a_k = \lim_{n\to\infty} \big(a_1 a_2 \cdots a_n\big),$$

provided that the limit exists. We now put $z = \pi/2$ into (5.4):

$$\frac{2}{\pi} = \cos\left(\frac{\pi}{2^2}\right) \cdot \cos\left(\frac{\pi}{2^3}\right) \cdot \cos\left(\frac{\pi}{2^4}\right) \cdot \cos\left(\frac{\pi}{2^5}\right) \cdots = \prod_{n=1}^{\infty} \cos\left(\frac{\pi}{2^{n+1}}\right)$$

We now just have to obtain formulas for $\cos\left(\frac{\pi}{2^n}\right)$. To do so, note that for any $0 \le \theta \le \pi$, we have

$$\cos\left(\frac{\theta}{2}\right) = \sqrt{\frac{1}{2} + \frac{1}{2}\cos\theta}.$$

(This follows from the double angle formula: $2\cos^2(2z) = 1 + \cos z$.) Thus,

$$\cos\left(\frac{\theta}{2^2}\right) = \sqrt{\frac{1}{2} + \frac{1}{2}\cos\left(\frac{\theta}{2}\right)} = \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\cos\theta}},$$

Continuing this process (slang for "it can be shown by induction"), we see that

$$\cos\left(\frac{\theta}{2^n}\right) = \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\cos\theta}}}}, \tag{5.5}$$

where there are n square roots here. Therefore, putting $\theta = \pi/2$ we obtain

$$\cos\left(\frac{\pi}{2^{n+1}}\right) = \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}}}}},$$

where there are n square roots here. In conclusion, we have shown that

$$\frac{2}{\pi} = \prod_{n=1}^{\infty} \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}}}}},$$

where there are n square roots in the n-th factor of the infinite product; or, writing out the infinite product, we have

$$\boxed{\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots .} \tag{5.6}$$

This formula was given by Viète in 1593.

5.1.2. Expansion of sine I. Our first proof of Euler's infinite product for sine is based on a neat identity involving tangents that we'll present in Lemma 5.1 below. To begin we first write, for $z \in \mathbb{C}$,

$$\sin z = \frac{1}{2i}\Big(e^{iz} - e^{-iz}\Big) = \lim_{n\to\infty} \frac{1}{2i}\left\{\left(1 + \frac{iz}{n}\right)^n - \left(1 - \frac{iz}{n}\right)^n\right\} = \lim_{n\to\infty} F_n(z), \tag{5.7}$$

where F_n is the polynomial of degree n in z given by

$$F_n(z) = \frac{1}{2i}\left\{\left(1 + \frac{iz}{n}\right)^n - \left(1 - \frac{iz}{n}\right)^n\right\}. \tag{5.8}$$

In the following lemma, we write $F_n(z)$ in terms of tangents.

LEMMA 5.1. *If $n = 2m + 1$ with $m \in \mathbb{N}$, then we can write*

$$F_n(z) = z \prod_{k=1}^{m} \left(1 - \frac{z^2}{n^2 \tan^2(k\pi/n)}\right).$$

PROOF. Observe that setting $z = n \tan\theta$, we have

$$\begin{aligned} 1 + \frac{iz}{n} = 1 + i\tan\theta = 1 + i\frac{\sin\theta}{\cos\theta} &= \frac{1}{\cos\theta}\big(\cos\theta + i\sin\theta\big) \\ &= \sec\theta\, e^{i\theta}, \end{aligned}$$

and similarly, $1 - iz/n = \sec\theta\, e^{-i\theta}$. Thus,

$$\begin{aligned} F_n(n\tan\theta) &= \frac{1}{2i}\left\{\left(1 + \frac{iz}{n}\right)^n - \left(1 - \frac{iz}{n}\right)^n\right\}\Bigg|_{z = n\tan\theta} \\ &= \frac{1}{2i}\sec^n\theta\Big(e^{in\theta} - e^{-in\theta}\Big), \end{aligned}$$

or, since $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$, we have

$$F_n(n\tan\theta) = \sec^n\theta\, \sin(n\theta).$$

The sine function vanishes at integer multiples of π, so it follows that $F_n(n\tan\theta) = 0$ where $n\theta = k\pi$ for all integers k, that is, for $\theta = k\pi/n$ for all $k \in \mathbb{Z}$. Thus, $F_n(z_k) = 0$ for

$$z_k = n\tan\left(\frac{k\pi}{n}\right) = n\tan\Big(\frac{k\pi}{2m+1}\Big),$$

where we recall that $n = 2m + 1$. Since $\tan\theta$ is strictly increasing on the interval $(-\pi/2, \pi/2)$, it follows that

$$z_{-m} < z_{-m+1} < \cdots < z_{-1} < z_0 < z_1 < \cdots < z_{m-1} < z_m;$$

moreover, since tangent is an odd function, we have $z_{-k} = -z_k$ for each k. In particular, we have found $2m + 1 = n$ distinct roots of $F_n(z)$, so as a consequence of the fundamental theorem of algebra, we can write $F_n(z)$ as a constant times

$$\begin{aligned} &(z - z_0) \cdot \prod_{k=1}^{m} \Big\{(z - z_k)\cdot(z - z_{-k})\Big\} \\ &\qquad = z \cdot \prod_{k=1}^{m} \Big\{(z - z_k)\cdot(z + z_k)\Big\} \quad (\text{since } z_{-k} = -z_k) \\ &\qquad = z \prod_{k=1}^{m} (z^2 - z_k^2) \\ &\qquad = z \prod_{k=1}^{m} \left(z^2 - n^2\tan^2\left(\frac{k\pi}{n}\right)\right). \end{aligned}$$

Factoring out the $-n^2\tan^2\left(\frac{k\pi}{n}\right)$ terms and gathering them all into one constant we can conclude that

$$F_n(z) = a\, z \prod_{k=1}^{m} \left(1 - \frac{z^2}{n^2\tan^2(k\pi/n)}\right),$$

for some constant a. Multiplying out the terms in the formula (5.8), we see that $F_n(z) = z$ plus higher powers of z. This implies that $a = 1$ and completes the proof of the lemma. □

Using this lemma we can give a formal[5] proof of Euler's sine expansion. From (5.7) and Lemma 5.1 we know that for any $x \in \mathbb{R}$,

$$\frac{\sin x}{x} = \lim_{n\to\infty} \frac{F_n(x)}{x} = \lim_{n\to\infty} \prod_{k=1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right), \tag{5.9}$$

where in the limit we restrict n to *odd* natural numbers. Thus, writing $n = 2m+1$ the limit in (5.9) really means

$$\frac{\sin x}{x} = \lim_{m\to\infty} \prod_{k=1}^{m} \left(1 - \frac{x^2}{(2m+1)^2 \tan^2(k\pi/(2m+1))}\right),$$

but we prefer the simpler form in (5.9) with the understanding that n is odd in (5.9). We now take $n \to \infty$ in this expression. Now,

$$\begin{aligned}\lim_{n\to\infty} n^2 \tan^2(k\pi/n) &= \lim_{n\to\infty} n^2 \frac{\sin^2(k\pi/n)}{\cos^2(k\pi/n)}\\ &= \lim_{n\to\infty} \left(\frac{\sin(k\pi/n)}{1/n} \cdot \frac{1}{\cos(k\pi/n)}\right)^2\\ &= \lim_{n\to\infty} \left((k\pi)^2 \cdot \frac{\sin(k\pi/n)}{k\pi/n} \cdot \frac{1}{\cos(k\pi/n)}\right)^2\\ &= k^2\pi^2,\end{aligned}$$

where we used that $\lim_{z\to 0} \frac{\sin z}{z} = 1$ and $\cos(0) = 1$. Hence,

$$\lim_{n\to\infty} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) = \left(1 - \frac{x^2}{k^2\pi^2}\right), \tag{5.10}$$

thus, formally evaluating the limit in (5.9), we see that

$$\begin{aligned}\frac{\sin x}{x} &= \lim_{n\to\infty} \prod_{k=1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right)\\ &= \prod_{k=1}^{\infty} \lim_{n\to\infty} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right)\\ &= \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{k^2\pi^2}\right),\end{aligned}$$

which is Euler's result. Unfortunately, there is one issue with this argument; it occurs in switching the limit with the product:

$$\lim_{n\to\infty} \prod_{k=1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) = \prod_{k=1}^{\infty} \lim_{n\to\infty} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right). \tag{5.11}$$

See Problem 2 for an example where such an interchange leads to a wrong answer. In Section 7.3 of Chapter 7 we'll learn a Tannery's theorem for infinite products,

[5] "Formal" in mathematics usually refers to "having the form or appearance without the substance or essence," which is the 5-th entry for "formal" in Webster's 1828 dictionary. This is very different to the common use of "formal": "according to form; agreeable to established mode; regular; methodical," which is the first entry in Webster's 1828 dictionary. Elaborating on the mathematicians use of "formal," it means something like "a symbolic manipulation or expression presented without paying attention to correctness".

from which we can easily deduce that (5.11) does indeed hold. However, we'll leave Tannery's theorem for products until Chapter 7 because we can easily justify (5.11) in a very elementary (although a little long-winded) way, which we do in the following theorem.

Theorem 5.2 (**Euler's theorem**). *For any $x \in \mathbb{R}$ we have*

$$\sin x = x \prod_{k=1}^{\infty} \Big(1 - \frac{x^2}{\pi^2 k^2}\Big).$$

Proof. We just have to verify the formula (5.11), which in view of (5.10) is equivalent to the equality

$$\lim_{n\to\infty} p_n = \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{k^2\pi^2}\right),$$

where

$$p_n = \prod_{k=1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right).$$

The limit in the case $x = 0$ is easily checked to hold so we can (and henceforth do) fix $x \neq 0$.

Step 1: We begin by finding some nice estimates on the quotient $\frac{x^2}{n^2\tan^2(k\pi/n)}$. In Lemma 4.55 back in Section 4.10, we proved that

$$\theta < \tan\theta, \qquad \text{for } 0 < \theta < \pi/2.$$

In particular, if $n \in \mathbb{N}$ is odd and $1 \leq k \leq \frac{n-1}{2}$, then

$$\frac{k\pi}{n} < \frac{n-1}{2} \cdot \frac{\pi}{n} < \frac{\pi}{2},$$

so

$$\frac{x^2}{n^2\tan^2(k\pi/n)} < \frac{x^2}{n^2(k\pi)^2/n^2} = \frac{x^2}{k^2\pi^2}. \tag{5.12}$$

Step 2: We now break up p_n in a nice way. Choose $m < \frac{n-1}{2}$ and let us break up the product p_n from $k = 1$ to m and then from m to $\frac{n-1}{2}$:

$$p_n = \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2\tan^2(k\pi/n)}\right) \prod_{k=m+1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2\tan^2(k\pi/n)}\right). \tag{5.13}$$

We shall use (5.12) to find estimates on the second product in (5.13). Choose m and n large enough such that $\frac{x^2}{m^2\pi^2} < 1$. Then it follows that from (5.12) that

$$\frac{x^2}{n^2\tan^2(k\pi/n)} < \frac{x^2}{k^2\pi^2} < 1 \quad \text{for} \quad k = m+1, m+2, \ldots, \frac{n-1}{2}.$$

In particular,

$$0 < \left(1 - \frac{x^2}{n^2\tan^2(k\pi/n)}\right) < 1 \quad \text{for} \quad k = m+1, m+2, \ldots, \frac{n-1}{2}.$$

Hence,

$$0 < \prod_{k=m+1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2\tan^2(k\pi/n)}\right) < 1$$

and therefore, in view of (5.13), we have

$$p_n \leq \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right).$$

In Problem 3 you will prove that for any nonnegative real numbers $a_1, a_2, \ldots, a_p \geq 0$, we have

$$(5.14) \qquad (1 - a_1)(1 - a_2) \cdots (1 - a_p) \geq 1 - (a_1 + a_2 + \cdots + a_p).$$

Using this inequality it follows that

$$\prod_{k=m+1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) \geq 1 - \sum_{k=m+1}^{\frac{n-1}{2}} \frac{x^2}{n^2 \tan^2(k\pi/n)}.$$

By (5.12) we have

$$\sum_{k=m+1}^{\frac{n-1}{2}} \frac{x^2}{n^2 \tan^2(k\pi/n)} \leq \frac{x^2}{\pi^2} \sum_{k=m+1}^{\frac{n-1}{2}} \frac{1}{k^2} \leq s_m,$$

where

$$s_m = \frac{x^2}{\pi^2} \sum_{k=m+1}^{\infty} \frac{1}{k^2}.$$

Thus,

$$1 - \sum_{k=m+1}^{\frac{n-1}{2}} \frac{x^2}{n^2 \tan^2(k\pi/n)} \geq 1 - s_m,$$

and hence,

$$\prod_{k=m+1}^{\frac{n-1}{2}} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) \geq 1 - s_m.$$

Therefore, in view of the expression (5.13) for p_n, we have

$$p_n \geq (1 - s_m) \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right).$$

To summarize, we have shown that

$$(5.15) \qquad (1 - s_m) \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) \leq p_n \leq \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right),$$

Step 3: Using (5.15) we can now finish the proof. Indeed, from (5.9) we know that

$$\lim_{n \to \infty} p_n(x) = \frac{\sin x}{x},$$

and recalling the limit (5.10), by the algebra of limits we have

$$\lim_{n \to \infty} \prod_{k=1}^{m} \left(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\right) = \prod_{k=1}^{m} \left(1 - \frac{x^2}{k^2 \pi^2}\right),$$

since the product $\prod_{k=1}^{m}$ is a finite product. Thus, taking $n \to \infty$ (5.15) we obtain

$$(1 - s_m) \prod_{k=1}^{m} \left(1 - \frac{x^2}{k^2 \pi^2}\right) \leq \frac{\sin x}{x} \leq \prod_{k=1}^{m} \left(1 - \frac{x^2}{k^2 \pi^2}\right),$$

or after rearrangement,

$$-s_m \prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right) \;\le\; \frac{\sin x}{x} \;-\; \prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right) \le 0.$$

Taking absolute values, we get

$$\left|\frac{\sin x}{x} - \prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right| \le |s_m| \left|\prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right|.$$

Our goal is to take $m \to \infty$ here, but before doing so we need the following estimate on the right-hand side:

$$\begin{aligned}\left|\prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right| &\le \prod_{k=1}^{m}\left(1+\frac{x^2}{k^2\,\pi^2}\right)\\ &\le \prod_{k=1}^{m} e^{\frac{x^2}{k^2\,\pi^2}} \quad (\text{since } 1+t \le e^t \text{ for any } t \in \mathbb{R})\\ &= e^{\sum_{k=1}^{m}\frac{x^2}{k^2\,\pi^2}} \quad (\text{since } e^a \cdot e^b = e^{a+b})\\ &\le e^L,\end{aligned}$$

where $L = \sum_{k=1}^{\infty}\frac{x^2}{k^2\,\pi^2}$, a finite constant the exact value of which is not important. (Note that $\sum_{k=1}^{\infty}\frac{1}{k^2}$ converges by the p-test with $p = 2$.) Thus,

$$\left|\prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right| \le e^L,$$

and so,

$$\left|\frac{\sin x}{x} - \prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right| \le |s_m| e^L. \tag{5.16}$$

Recalling that $s_m = \frac{x^2}{\pi^2}\sum_{k=m+1}^{\infty}\frac{1}{k^2}$ and $\sum_{k=1}^{\infty}\frac{1}{k^2}$ converges (the p-test with $p = 2$), by the Cauchy Criterion for series we know that $\lim_{m\to\infty} s_m = 0$. Thus, it follows from (5.16) that

$$\left|\frac{\sin x}{x} - \prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right)\right|$$

can be made as small as we desire by taking m larger and larger. This, by definition of limit, means that

$$\frac{\sin x}{x} = \lim_{m\to\infty}\prod_{k=1}^{m}\left(1-\frac{x^2}{k^2\,\pi^2}\right),$$

which proves our result. □

We remark that Euler's sine expansion also holds for all complex $z \in \mathbb{C}$ (and not just real $x \in \mathbb{R}$), but we'll wait for Section 7.3 of Chapter 7 for the proof of the complex version.

5.1.3. Wallis' formulas. As an application of Euler's sine expansion, we can derive John Wallis' (1616–1703) formulas for π.

THEOREM 5.3 (**Wallis' formulas**). *We have*

$$\boxed{\begin{aligned}\frac{\pi}{2} &= \prod_{n=1}^{\infty} \frac{2n}{2n-1}\cdot\frac{2n}{2n+1} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots,\\ \sqrt{\pi} &= \lim_{n\to\infty}\frac{1}{\sqrt{n}}\prod_{k=1}^{n}\frac{2k}{2k-1} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{2}{1}\cdot\frac{4}{3}\cdot\frac{6}{5}\cdots\frac{2n}{2n-1}.\end{aligned}}$$

PROOF. To obtain the first formula, we set $x = \pi/2$ in Euler's infinite product expansion for sine:

$$\sin x = x\prod_{n=1}^{\infty}\Big(1 - \frac{x^2}{\pi^2 n^2}\Big) \quad\Longrightarrow\quad 1 = \frac{\pi}{2}\prod_{n=1}^{\infty}\Big(1 - \frac{1}{2^2 n^2}\Big).$$

Since $1 - \frac{1}{2^2n^2} = \frac{2^2n^2-1}{2^2n^2} = \frac{(2n-1)(2n+1)}{(2n)(2n)}$, we see that

$$\frac{2}{\pi} = \prod_{n=1}^{\infty}\frac{2n-1}{2n}\cdot\frac{2n+1}{2n}.$$

Now taking reciprocals of both sides (you are encouraged to verify that the reciprocal of an infinite product is the product of the reciprocals) we get Wallis' first formula. To obtain the second formula, we write the first formula as

$$\frac{\pi}{2} = \lim_{n\to\infty}\left\{\Big(\frac{2}{1}\Big)^2\cdot\Big(\frac{4}{3}\Big)^2\cdots\Big(\frac{2n}{2n-1}\Big)^2\cdot\frac{1}{2n+1}\right\}.$$

Then taking square roots we obtain

$$\sqrt{\pi} = \lim_{n\to\infty}\sqrt{\frac{2}{2n+1}}\prod_{k=1}^{n}\frac{2k}{2k-1} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\frac{1}{\sqrt{1+1/2n}}\prod_{k=1}^{n}\frac{2k}{2k-1}.$$

Using that $1/\sqrt{1+1/2n} \to 1$ as $n \to \infty$ completes our proof. □

We prove prove a beautiful expression for π due to Sondow [**217**] (which I found on Weisstein's website [**241**]). To present this formula, we first manipulate Wallis' first formula to

$$\frac{\pi}{2} = \prod_{n=1}^{\infty}\frac{2n}{2n-1}\cdot\frac{2n}{2n+1} = \prod_{n=1}^{\infty}\frac{4n^2}{4n^2-1} = \prod_{n=1}^{\infty}\left(1 + \frac{1}{4n^2-1}\right).$$

Second, using partial fractions we observe that

$$\sum_{n=1}^{\infty}\frac{1}{4n^2-1} = \frac{1}{2}\sum_{n=1}^{\infty}\left(\frac{1}{2n-1} - \frac{1}{2n+1}\right) = \frac{1}{2}\cdot 1 = \frac{1}{2},$$

since the sum telescopes (see e.g. the telescoping series theorem — Theorem 3.24). Dividing these two formulas, we get

$$\boxed{\pi = \frac{\displaystyle\prod_{n=1}^{\infty}\left(1 + \frac{1}{4n^2-1}\right)}{\displaystyle\sum_{n=1}^{\infty}\frac{1}{4n^2-1}},}$$

quite astonishing!

Exercises 5.1.

1. Here are some Viète-Wallis products from [**175, 178**].
   (i) From the formulas (5.3), (5.5), and Euler's sine expansion, prove that for any $x \in \mathbb{R}$ and $p \in \mathbb{N}$ we have

$$\frac{\sin x}{x} = \prod_{k=1}^{p} \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\cos x}}}} \cdot \prod_{n=1}^{\infty} \Big(\frac{2^p\pi n - \theta}{2^p\pi n} \cdot \frac{2^p\pi n + \theta}{2^p\pi n}\Big),$$

   where there are k square roots in the k-th factor of the product $\prod_{k=1}^{p}$.
   (ii) Setting $x = \pi/2$ in (i), show that

$$\frac{2}{\pi} = \prod_{k=1}^{p} \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}}}}} \cdot \prod_{n=1}^{\infty} \Big(\frac{2^{p+1}n - 1}{2^{p+1}n} \cdot \frac{2^{p+1}n + 1}{2^{p+1}n}\Big),$$

   where there are k square roots in the k-th factor of the product $\prod_{k=1}^{p}$.
   (iii) Setting $x = \pi/6$ in (i), show that

$$\frac{3}{\pi} = \prod_{k=1}^{p} \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \cdots + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\Big(\frac{\sqrt{3}}{2}\Big)}}}} \cdot \prod_{n=1}^{\infty} \Big(\frac{3 \cdot 2^{p+1}n - 1}{3 \cdot 2^{p+1}n} \cdot \frac{3 \cdot 2^{p+1}n + 1}{3 \cdot 2^{p+1}n}\Big),$$

   where there are k square roots in the k-th factor of the product $\prod_{k=1}^{p}$.
   (iv) Experiment with two other values of x to derive other Viète-Wallis-type formulas.
2. Suppose for each $n \in \mathbb{N}$ we are given a finite product

$$\prod_{k=1}^{a_n} f_k(n),$$

   where $f_k(n)$ is an expression involving k, n and $a_n \in \mathbb{N}$ is such that $\lim_{n\to\infty} a_n = \infty$. For example, in (5.11) we have $a_n = \frac{n-1}{2}$ and $f_k(n) = \Big(1 - \frac{x^2}{n^2 \tan^2(k\pi/n)}\Big)$; then (5.11) claims that for this example we have

$$\lim_{n\to\infty} \prod_{k=1}^{a_n} f_k(n) = \prod_{k=1}^{\infty} \lim_{n\to\infty} f_k(n). \tag{5.17}$$

   However, this equality is not always true. Indeed, prove that (5.17) is false for the example $a_n = n$ and $f_k(n) = 1 + \frac{1}{n}$.
3. Prove (5.14) using induction on p.
4. Prove the following splendid formula:

$$\boxed{\sqrt{\pi} = \lim_{n\to\infty} \frac{(n!)^2\, 2^{2n}}{(2n)!\,\sqrt{n}}.}$$

   Suggestion: Wallis' formula is hidden here.
5. (cf. [**22**]) In this problem we give an elementary proof of the following interesting identity: For any n that is a power of 2 and for any $x \in \mathbb{R}$ we have

$$\sin x = n \sin\Big(\frac{x}{n}\Big) \cos\Big(\frac{x}{n}\Big) \prod_{k=1}^{\frac{n}{2}-1} \Big(1 - \frac{\sin^2(x/n)}{\sin^2(k\pi/n)}\Big). \tag{5.18}$$

   (i) Prove that for any $x \in \mathbb{R}$,

$$\sin x = 2 \sin\Big(\frac{x}{2}\Big) \sin\Big(\frac{\pi + x}{2}\Big).$$

(ii) Show that for n equal to any power of 2, we have

$$\sin x = 2^n \sin\Big(\frac{x}{n}\Big) \sin\Big(\frac{\pi + x}{n}\Big) \sin\Big(\frac{2\pi + x}{n}\Big) \cdots \\ \cdots \sin\Big(\frac{(n-2)\pi + x}{n}\Big) \sin\Big(\frac{(n-1)\pi + x}{n}\Big);$$

note that if $n = 2^1$ we get the formula in (i).

(iii) Show that the formula in (ii) can be written as

$$\sin x = 2^n \sin\Big(\frac{x}{n}\Big) \sin\Big(\frac{\frac{n}{2}\pi + x}{n}\Big) \prod_{1 \le k < \frac{n}{2}} \sin\Big(\frac{k\pi + x}{n}\Big) \sin\Big(\frac{k\pi - x}{n}\Big).$$

(iv) Prove the identity $\sin(\theta + \varphi)\, \sin(\theta - \varphi) = \sin^2 \theta - \sin^2 \varphi$ and use this to conclude that the formula in (iii) equals

$$\sin x = 2^n \sin\Big(\frac{x}{n}\Big) \cos\Big(\frac{x}{n}\Big) \prod_{1 \le k < \frac{n}{2}} \bigg(\sin^2\Big(\frac{k\pi}{n}\Big) - \sin^2\Big(\frac{x}{n}\Big) \bigg).$$

(v) By considering what happens as $x \to 0$ in the formula in (iv), prove that for n a power of 2, we have

$$n = 2^n \prod_{1 \le k < \frac{n}{2}} \bigg(\sin^2\Big(\frac{k\pi}{n}\Big) \bigg).$$

Now prove (5.18).

6. (**Expansion of sine II**) We give a second proof of Euler's sine expansion.

(i) Show that taking $n \to \infty$ on both sides of the identity (5.18) from the previous problem gives a *formal* proof of Euler's sine expansion.

(ii) Now using the identity (5.18) and following the ideas found in the proof of Theorem 5.2, give another rigorous proof of Euler's sine expansion.

5.2. ★ Euler, Gregory, Leibniz, and Madhava

In this section we present two beautiful formulas involving π: Euler's formula for $\pi^2/6$ and the Gregory-Leibniz-Madhava formula for $\pi/4$. The simplest proofs I know of these formulas are taken from the article by Hofbauer [**102**] and are completely "elementary" in the sense that they involve nothing involving derivatives or integrals ... just a little bit of trigonometric identities and then a dab of some inequalities (or Tannery's theorem if you prefer) to finish them off. However, before presenting Hofbauer's proofs we present (basically) Euler's original (third) proof of his solution to the Basel problem.

5.2.1. Proof I of Euler's formula for $\pi^2/6$.

In 1644, the Italian mathematician Pietro Mengoli (1625–1686) posed the question: What's the value of the sum

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots ?$$

This problem was made popular through Jacob (Jacques) Bernoulli (1654–1705) when he wrote about it in 1689 and was solved by Leonhard Euler (1707–1783) in 1735. Bernoulli was so baffled by the unknown value of the series that he wrote

> *If somebody should succeed in finding what till now withstood our efforts and communicate it to us, we shall be much obliged to him.* [**47**, p. 73], [**252**, p. 345].

Before Euler's solution to this request, known as the *Basel problem* (Bernoulli lived in Basel, Switzerland), this problem eluded many of the great mathematicians of that day: In 1742, Euler wrote

> *Jacob Bernoulli does mention those series, but confesses that, in spite of all his efforts, he could not get through, so that Joh. Bernoulli, de Moivre and Stirling, great authorities in such matters, were highly surprised when I told them that I had found the sum of $\zeta(2)$, and even of $\zeta(n)$ for n even.* [**237**, pp. 262-63].

(We shall consider $\zeta(n)$ for n even in the next section.) Needless to say, it shocked the mathematical community when Euler found the sum to be $\pi^2/6$; in the introduction to his famous 1735 paper *De summis serierum reciprocarum* (On the sums of series of reciprocals) where he first proves that $\zeta(2) = \pi^2/6$, Euler writes:

> *So much work has been done on the series $\zeta(n)$ that it seems hardly likely that anything new about them may still turn up ... I, too, in spite of repeated efforts, could achieve nothing more than approximate values for their sums ... Now, however, quite unexpectedly, I have found an elegant formula for $\zeta(2)$, depending upon the quadrature of the circle [i.e., upon π]* [**237**, p. 261].

For more on various solutions to the Basel problem, see [**109**], [**49**], [**195**], and for more on Euler, see [**11**], [**119**]. On the side is a picture of a Swiss Franc banknote honoring Euler.

We already saw Euler's original argument in the introduction to this chapter; we shall now make his argument rigorous. First, we claim that for any nonnegative real numbers $a_1, a_2, \ldots, a_n \geq 0$, we have

$$1 - \sum_{k=1}^{n} a_k \leq \prod_{k=1}^{n} (1 - a_k) \leq 1 - \sum_{k=1}^{n} a_k + \sum_{1 \leq i < j \leq n} a_i \, a_j . \tag{5.19}$$

You will prove these inequalities in Problem 1. Applying these inequalities to $\prod_{k=1}^{n} \left(1 - \frac{x^2}{k^2\pi^2}\right)$, we obtain

$$1 - \sum_{k=1}^{n} \frac{x^2}{k^2\pi^2} \leq \prod_{k=1}^{n} \left(1 - \frac{x^2}{k^2\pi^2}\right) \leq 1 - \sum_{k=1}^{n} \frac{x^2}{k^2\pi^2} + \sum_{1 \leq i < j \leq n} \frac{x^2}{i^2\pi^2} \frac{x^2}{j^2\pi^2} .$$

After some slight simplifications we can write this as

$$1 - \frac{x^2}{\pi^2} \sum_{k=1}^{n} \frac{1}{k^2} \leq \prod_{k=1}^{n} \left(1 - \frac{x^2}{k^2\pi^2}\right) \leq 1 - \frac{x^2}{\pi^2} \sum_{k=1}^{n} \frac{1}{k^2} + \frac{x^4}{\pi^4} \sum_{1 \leq i < j \leq n} \frac{1}{i^2 \, j^2} . \tag{5.20}$$

Let us put

$$\zeta_n(2) = \sum_{k=1}^{n} \frac{1}{k^2} \quad \text{and} \quad \zeta_n(4) = \sum_{k=1}^{n} \frac{1}{k^4} ,$$

and observe that

$$\begin{aligned}\zeta_n(2)^2 = \left(\sum_{i=1}^n \frac{1}{i^2}\right)\left(\sum_{j=1}^n \frac{1}{j^2}\right) &= \sum_{i,j=1}^n \frac{1}{i^2\,j^2}\\ &= \zeta_n(4) + 2\sum_{1\le i<j\le n} \frac{1}{i^2\,j^2}.\end{aligned}$$

Thus, (5.20) can be written as

$$1 - \frac{x^2}{\pi^2}\zeta_n(2) \le \prod_{k=1}^n \left(1 - \frac{x^2}{k^2\pi^2}\right) \le 1 - \frac{x^2}{\pi^2}\zeta_n(2) + \frac{x^4}{\pi^4}\frac{\zeta_n(2)^2 - \zeta_n(4)}{2}.$$

We remark that the exact coefficient of x^4 on the right is not important, but it might be helpful if you try Problem 7. Taking $n \to \infty$ and using that $\zeta_n(2) \to \zeta(2)$, $\prod_{k=1}^n \left(1 - \frac{x^2}{k^2\pi^2}\right) \to \frac{\sin x}{x}$, and that $\zeta_n(4) \to \zeta(4)$, we obtain

$$1 - \frac{x^2}{\pi^2}\zeta(2) \le \frac{\sin x}{x} \le 1 - \frac{x^2}{\pi^2}\zeta(2) + \frac{x^4}{\pi^4}\frac{\zeta(2)^2 - \zeta(4)}{2}.$$

Replacing $\frac{\sin x}{x}$ by its power series expansion we see that

$$1 - \frac{x^2}{\pi^2}\zeta(2) \le 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots \le 1 - \frac{x^2}{\pi^2}\zeta(2) + \frac{x^4}{\pi^4}\frac{\zeta(2)^2 - \zeta(4)}{2}.$$

Now subtracting 1 from everything and dividing by x^2, we get

$$-\frac{1}{\pi^2}\zeta(2) \le -\frac{1}{3!} + \frac{x^2}{5!} - \frac{x^4}{7!} + \cdots \le -\frac{1}{\pi^2}\zeta(2) + \frac{x^2}{\pi^4}\frac{\zeta(2)^2 - \zeta(4)}{2}.$$

Finally, putting $x = 0$ we conclude that

$$-\frac{1}{\pi^2}\zeta(2) \le -\frac{1}{3!} \le -\frac{1}{\pi^2}\zeta(2).$$

This implies that $\zeta(2) = \frac{\pi^2}{6}$, exactly as Euler stated.

5.2.2. Proof II of Euler's formula for $\pi^2/6$. Follow Hofbauer [**102**] we give our second proof of Euler's formula for $\pi^2/6$. We begin with the identity, valid for noninteger $z \in \mathbb{C}$,

$$\frac{1}{\sin^2 z} = \frac{1}{4\sin^2\frac{z}{2}\cos^2\frac{z}{2}} = \frac{1}{4}\left(\frac{1}{\sin^2\frac{z}{2}} + \frac{1}{\cos^2\frac{z}{2}}\right) = \frac{1}{4}\left(\frac{1}{\sin^2\frac{z}{2}} + \frac{1}{\sin^2\left(\frac{\pi - z}{2}\right)}\right),$$

where at the last step we used that $\cos(z) = \sin(\frac{\pi}{2} - z)$. Replacing z with πz, we get for noninteger z,

$$\frac{1}{\sin^2 \pi z} = \frac{1}{4}\left(\frac{1}{\sin^2\frac{z\pi}{2}} + \frac{1}{\sin^2\left(\frac{(1-z)\pi}{2}\right)}\right). \tag{5.21}$$

In particular, setting $z = 1/2$, we obtain

$$1 = \frac{1}{4}\left(\frac{1}{\sin^2\frac{\pi}{2^2}} + \frac{1}{\sin^2\frac{\pi}{2^2}}\right) = \frac{2}{4}\cdot\frac{1}{\sin^2\frac{\pi}{2^2}}.$$

Applying (5.21) (with $z = 1/2^2$) to the right-hand side of this equation gives

$$1 = \frac{2}{4^2}\left(\frac{1}{\sin^2\frac{\pi}{2^3}} + \frac{1}{\sin^2\frac{3\pi}{2^3}}\right) = \frac{2}{4^2}\sum_{k=0}^{1}\frac{1}{\sin^2\frac{(2k+1)\pi}{2^3}}.$$

Applying (5.21) to each term $\frac{1}{\sin^2\frac{\pi}{2^3}}$ and $\frac{1}{\sin^2\frac{3\pi}{2^3}}$ gives

$$\begin{aligned}1 &= \frac{2}{4^2}\left(\frac{1}{4}\left[\frac{1}{\sin^2\frac{\pi}{2^4}} + \frac{1}{\sin^2\frac{7\pi}{2^4}}\right] + \frac{1}{4}\left[\frac{1}{\sin^2\frac{3\pi}{2^4}} + \frac{1}{\sin^2\frac{5\pi}{2^4}}\right]\right)\\ &= \frac{2}{4^3}\left(\frac{1}{\sin^2\frac{\pi}{2^4}} + \frac{1}{\sin^2\frac{3\pi}{2^4}} + \frac{1}{\sin^2\frac{5\pi}{2^4}} + \frac{1}{\sin^2\frac{7\pi}{2^4}}\right)\\ &= \frac{2}{4^3}\sum_{k=0}^{2}\frac{1}{\sin^2\frac{(2k+1)\pi}{2^4}}.\end{aligned}$$

Repeatedly applying (5.21) (slang for "use induction"), we arrive at the following.

LEMMA 5.4. *For any $n \in \mathbb{N}$, we have*

$$1 = \frac{2}{4^n}\sum_{k=0}^{2^{n-1}-1}\frac{1}{\sin^2\frac{(2k+1)\pi}{2^{n+1}}}.$$

To establish Euler's formula, we need one more lemma.

LEMMA 5.5. *For $0 < x < \pi/2$, we have*

$$-1 + \frac{1}{\sin^2 x} < \frac{1}{x^2} < \frac{1}{\sin^2 x}.$$

PROOF. Taking reciprocals in the formula from Lemma 4.55: For $0 < x < \pi/2$,

$$\sin x < x < \tan x,$$

we get $\cot^2 x < x^{-2} < \sin^{-2} x$. Since $\cot^2 x = \cos^2 x/\sin^2 x = \sin^{-2} x - 1$, we conclude that

$$\frac{1}{\sin^2 x} > \frac{1}{x^2} > -1 + \frac{1}{\sin^2 x}, \quad 0 < x < \frac{\pi}{2},$$

which proves the lemma. □

Now, observe that for $0 \le k \le 2^{n-1} - 1$ we have

$$\frac{(2k+1)\pi}{2^{n+1}} \le \frac{(2(2^{n-1}-1)+1)\pi}{2^{n+1}} = \frac{(2^n-1)\pi}{2^{n+1}} < \frac{\pi}{2},$$

therefore using the identity

$$-1 + \frac{1}{\sin^2 x} < \frac{1}{x^2} < \frac{1}{\sin^2 x}, \quad 0 < x < \frac{\pi}{2}$$

we see that

$$-2^{n-1} + \sum_{k=0}^{2^{n-1}-1}\frac{1}{\sin^2\frac{(2k+1)\pi}{2^{n+1}}} < \sum_{k=0}^{2^{n-1}-1}\frac{1}{\left(\frac{(2k+1)\pi}{2^{n+1}}\right)^2} < \sum_{k=0}^{2^{n-1}-1}\frac{1}{\sin^2\frac{(2k+1)\pi}{2^{n+1}}}.$$

Multiplying both sides by $2/4^n = 2/2^{2n}$ and using Lemma 5.4, we get

$$-\frac{1}{2^n} + 1 < \frac{8}{\pi^2}\sum_{k=0}^{2^{n-1}-1}\frac{1}{(2k+1)^2} < 1.$$

Taking $n \to \infty$ and using the squeeze theorem, we conclude that

$$1 \le \frac{8}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} \le 1 \qquad \Longrightarrow \qquad \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8}.$$

Finally, summing over the even and odd numbers (see Problem 2a in Exercises 3.5), we have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} + \sum_{n=1}^{\infty} \frac{1}{(2n)^2} = \frac{\pi^2}{8} + \frac{1}{4} \sum_{n=1}^{\infty} \frac{1}{n^2} \tag{5.22}$$

$$\Longrightarrow \quad \frac{3}{4} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{8}.$$

and solving for $\sum_{n=1}^{\infty} 1/n^2$, we obtain Euler's formula:

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots .$$

5.2.3. Proof III of Euler's formula for $\pi^2/6$. In Proof II we established Euler's formula from Lemma 5.5. This time we apply Tannery's theorem. The idea is to write the identity in Lemma 5.4 in a form found in Tannery's theorem:

$$1 = \frac{2}{4^n} \sum_{k=0}^{2^{n-1}-1} \frac{1}{\sin^2 \frac{(2k+1)\pi}{2^{n+1}}} = \sum_{k=0}^{2^{n-1}-1} a_k(n), \tag{5.23}$$

where

$$a_k(n) = \frac{2}{4^n} \frac{1}{\sin^2 \frac{(2k+1)\pi}{2^{n+1}}}.$$

Let us verify the hypotheses of Tannery's theorem. First, since $\lim_{z\to 0} \frac{\sin z}{z} = 1$, we have

$$\lim_{n\to\infty} 2^{n+1} \sin \frac{(2k+1)\pi}{2^{n+1}} = (2k+1)\pi \cdot \lim_{n\to\infty} \frac{\sin \frac{(2k+1)\pi}{2^{n+1}}}{\frac{(2k+1)\pi}{2^{n+1}}} = (2k+1)\pi.$$

Therefore,

$$\begin{aligned} \lim_{n\to\infty} a_k(n) &= \lim_{n\to\infty} \frac{2}{4^n} \cdot \frac{1}{\sin^2 \frac{(2k+1)\pi}{2^{n+1}}} \\ &= \lim_{n\to\infty} 8 \cdot \frac{1}{\left(2^{n+1} \sin \frac{(2k+1)\pi}{2^{n+1}}\right)^2} = \frac{8}{\pi^2(2k+1)^2}. \end{aligned}$$

To verify the other hypothesis of Tannery's theorem we need the following lemma.

Lemma 5.6. *There exists a constant $c > 0$ such that for $0 \le x \le \pi/2$,*

$$c\,x \le \sin x.$$

Proof. Since $\lim_{z\to 0} \frac{\sin z}{z} = 1$, the function $f(x) = \sin x/x$ is a continuous function of x in $[0, \pi/2]$ where we define $f(0) := 1$. Observe that f is positive on $[0, \pi/2]$ because $f(0) = 1 > 0$ and $\sin x > 0$ for $0 < x \le \pi/2$. Therefore, by the max/min value theorem, $f(x) \ge f(b) > 0$ on $[0, \pi/2]$ for some $b \in [0, \pi/2]$. This proves that $c\,x \le \sin x$ on $[0, \pi/2]$ where $c = f(b) > 0$. □

Now, observe that for $0 \le k \le 2^{n-1} - 1$ we have

$$\frac{(2k+1)\pi}{2^{n+1}} \le \frac{(2(2^{n-1}-1)+1)\pi}{2^{n+1}} = \frac{(2^n-1)\pi}{2^{n+1}} < \frac{\pi}{2},$$

therefore by Lemma 5.6,

$$c \cdot \frac{(2k+1)\pi}{2^{n+1}} \le \sin \frac{(2k+1)\pi}{2^{n+1}} \quad \Longrightarrow \quad \frac{1}{\sin^2 \frac{(2k+1)\pi}{2^{n+1}}} \le \frac{4^{n+1}}{c^2\pi^2(2k+1)^2}.$$

Multiplying both sides by $2/4^n$, we obtain

$$\frac{2}{4^n} \cdot \frac{1}{\sin^2 \frac{(2k+1)\pi}{2^{n+1}}} \le \frac{8}{c^2\pi^2} \cdot \frac{1}{(2k+1)^2} =: M_k.$$

It follows that $|a_k(n)| \le M_k$ for all n, k. Moreover, since the sum $\sum_{k=0}^\infty M_k = \sum_{k=0}^\infty \frac{8}{c^2\pi^2} \cdot \frac{1}{(2k+1)^2}$ converges, we have verified the hypotheses of Tannery's theorem. Hence, taking $n \to \infty$ in (5.23), we get

$$\begin{aligned} 1 = \lim_{n\to\infty} \sum_{k=0}^{2^{n-1}-1} a_k(n) &= \sum_{k=0}^{\infty} \lim_{n\to\infty} a_k(n) \\ &= \sum_{k=0}^{\infty} \frac{8}{\pi^2(2k+1)^2} \quad \Longrightarrow \quad \frac{\pi^2}{8} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2}. \end{aligned}$$

Doing the even-odd trick as we did in (5.22), we know that this formula implies Euler's formula for $\pi^2/6$. See Problem 5 for Proof IV, a classic proof.

5.2.4. Proof I of Gregory-Leibniz-Madhava's formula for $\pi/4$. As the proof of Euler's formula was based on a trigonometric identity for sines (Lemma 5.4), the proof of Gregory-Leibniz-Madhava's formula:

$$\boxed{\frac{\pi}{4} = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{2n-1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots,}$$

also involves trigonometric identities, but for cotangents. Concerning Leibniz's discovery of this formula, Christian Huygens (1629–1695) wrote "that it would be a discovery always to be remembered among mathematicians" [**252**, p. 316]. In 1676, Isaac Newton (1642–1727) wrote

> *Leibniz's method for obtaining convergent series is certainly very elegant, and it would have sufficiently revealed the genius of its author, even if he had written nothing else.* [**226**, p. 130].

To prove the Gregory-Leibniz-Madhava formula we begin with the double angle formula

$$2\cot 2z = 2\frac{\cos 2z}{\sin 2z} = \frac{\cos^2 z - \sin^2 z}{\cos z \sin z} = \cot z - \tan z,$$

from which we see that

$$\cot 2z = \frac{1}{2}\Big(\cot z - \tan z\Big).$$

Since $\tan z = \cot(\pi/2 - z)$, we find that

$$\cot 2z = \frac{1}{2}\bigg(\cot z - \cot\Big(\frac{\pi}{2} - z\Big)\bigg).$$

Replacing z with $\pi z/2$, we get

$$\cot \pi z = \frac{1}{2}\bigg(\cot \frac{z\pi}{2} - \cot \frac{(1-z)\pi}{2} \bigg), \tag{5.24}$$

which is our fundamental equation. In particular, setting $z = 1/4$, we obtain

$$1 = \frac{1}{2}\bigg(\cot \frac{\pi}{4\cdot 2} - \cot \frac{3\pi}{4 \cdot 2} \bigg) = \frac{1}{2} \sum_{k=0}^{0} \bigg(\cot \frac{(4k+1)\pi}{2^3} - \cot \frac{(4k+3)\pi}{2^3} \bigg).$$

Applying (5.24) to each term $\cot \frac{\pi}{2^3}$ and $\cot \frac{3\pi}{2^3}$ gives

$$\begin{aligned} 1 &= \frac{1}{2} \left[\frac{1}{2}\bigg(\cot \frac{\pi}{2^4} - \cot \frac{7\pi}{2^4} \bigg) - \frac{1}{2}\bigg(\cot \frac{3\pi}{2^4} - \cot \frac{5\pi}{2^4} \bigg) \right] \\ &= \frac{1}{2^2} \left[\bigg(\cot \frac{\pi}{2^4} - \cot \frac{3\pi}{2^4} \bigg) + \bigg(\cot \frac{5\pi}{2^4} - \cot \frac{7\pi}{2^4} \bigg) \right] \\ &= \frac{1}{2^2} \sum_{k=0}^{1} \bigg(\cot \frac{(4k+1)\pi}{2^4} - \cot \frac{(4k+3)\pi}{2^4} \bigg). \end{aligned}$$

Repeatedly applying (5.24), one can prove that for any $n \in \mathbb{N}$, we have

$$1 = \frac{1}{2^n} \sum_{k=0}^{2^{n-1}-1} \bigg(\cot \frac{(4k+1)\pi}{2^{n+2}} - \cot \frac{(4k+3)\pi}{2^{n+2}} \bigg). \tag{5.25}$$

(The diligent reader will supply the details!) Since we know some nice properties of sine from Lemma 5.6 we write the right-hand side of this identity in terms of sine. To do so, observe that for any complex numbers z, w, not integer multiples of π, we have

$$\begin{aligned} \cot z - \cot w = \frac{\cos z}{\sin z} - \frac{\cos w}{\sin w} &= \frac{\sin w \cos z - \cos w \sin z}{\sin z \, \sin w} \\ &= \frac{\sin(w - z)}{\sin z \, \sin w}. \end{aligned}$$

Using this identity in (5.25), we obtain

$$1 = \frac{1}{2^n} \sum_{k=0}^{2^{n-1}-1} \frac{\sin \frac{\pi}{2^{n+1}}}{\sin \frac{(4k+1)\pi}{4\cdot 2^n} \cdot \sin \frac{(4k+3)\pi}{2^{n+2}}} = \sum_{k=0}^{2^{n-1}-1} a_k(n), \tag{5.26}$$

where

$$a_k(n) = \frac{1}{2^n} \frac{\sin \frac{\pi}{2^{n+1}}}{\sin \frac{(4k+1)\pi}{2^{n+2}} \cdot \sin \frac{(4k+3)\pi}{2^{n+2}}}.$$

The idea to derive Gregory-Leibniz-Madhava's formula is to take $n \to \infty$ in (5.26) and use Tannery's theorem. Let us verify the hypotheses of Tannery's theorem. First, to determine $\lim_{n\to\infty} a_k(n)$ we write

$$\begin{aligned} \frac{1}{2^n} \cdot \frac{\sin \frac{\pi}{2^{n+1}}}{\sin \frac{(4k+1)\pi}{4\cdot 2^n} \cdot \sin \frac{(4k+3)\pi}{4\cdot 2^n}} &= 2^3 \cdot \frac{2^{n+1}}{2^{n+2} \cdot 2^{n+2}} \cdot \frac{\sin \frac{\pi}{2^{n+1}}}{\sin \frac{(4k+1)\pi}{2^{n+2}} \cdot \sin \frac{(4k+3)\pi}{2^{n+2}}} \\ &= \frac{8}{\pi(4k+1)(4k+3)} \cdot \frac{\frac{2^{n+1}}{\pi} \sin \frac{\pi}{2^{n+1}}}{\left(\frac{2^{n+2}}{(4k+1)\pi} \sin \frac{(4k+1)\pi}{2^{n+2}} \right) \cdot \left(\frac{2^{n+2}}{(4k+3)\pi} \sin \frac{(4k+3)\pi}{2^{n+2}} \right)}. \end{aligned}$$

Therefore, since $\lim_{z\to 0}\frac{\sin z}{z} = 1$, we have

$$\lim_{n\to\infty} a_k(n) = \frac{8}{\pi(4k+1)(4k+3)}.$$

To verify the other hypothesis of Tannery's theorem we need the following lemma.

Lemma 5.7. *If $|z| \le 1$, then*

$$|\sin z| \le \frac{6}{5}|z|.$$

Proof. Observe that for $|z| \le 1$, we have $|z|^k \le |z|$ for any $k \in \mathbb{N}$, and

$$\begin{aligned}(2n+1)! &= (2\cdot 3)\cdot(4\cdot 5)\cdots(2n\cdot(2n+1))\\ &\ge (2\cdot 3)\cdot(2\cdot 3)\cdots(2\cdot 3) = (2\cdot 3)^n = 6^n.\end{aligned}$$

Thus,

$$\begin{aligned}|\sin z| = \left|\sum_{n=0}^\infty (-1)^n \frac{z^{2n+1}}{(2n+1)!}\right| &\le \sum_{n=0}^\infty \frac{|z|^{2n+1}}{(2n+1)!}\\ &\le \sum_{n=0}^\infty \frac{|z|}{6^n} = |z|\sum_{n=0}^\infty \frac{1}{6^n} = \frac{1}{1-(1/6)}|z| = \frac{6}{5}|z|.\end{aligned}$$

□

Since $0 \le \frac{\pi}{2^{n+1}} \le 1$ for $n \in \mathbb{N}$ (because $\pi < 4$), by this lemma we have

$$\sin\frac{\pi}{2^{n+1}} \le \frac{6}{5}\cdot\frac{\pi}{2^{n+1}}. \tag{5.27}$$

Observe that for $0 \le k \le 2^{n-1}-1$ and $0 \le \ell \le 4$, we have

$$\frac{(4k+\ell)\pi}{2^{n+2}} \le \frac{(4(2^{n-1}-1)+\ell)\pi}{2^{n+2}} = \frac{(2^{n+1}-4+\ell)\pi}{2^{n+2}} \le \frac{\pi}{2},$$

therefore by Lemma 5.6,

$$c\cdot\frac{(4k+\ell)\pi}{2^{n+2}} \le \sin\frac{(4k+\ell)\pi}{2^{n+2}} \quad\Longrightarrow\quad \frac{1}{\sin\frac{(4k+\ell)\pi}{2^{n+2}}} \le \frac{1}{c}\frac{2^{n+2}}{(4k+\ell)\pi}.$$

Combining this inequality with (5.27), we see that for $0 \le k \le 2^{n-1}-1$, we have

$$\begin{aligned}\frac{1}{2^n}\frac{\sin\frac{\pi}{2^{n+1}}}{\sin\frac{(4k+1)\pi}{2^{n+2}}\cdot\sin\frac{(4k+3)\pi}{2^{n+2}}} &\le \frac{1}{2^n}\cdot\left(\frac{6}{5}\cdot\frac{\pi}{2^{n+1}}\right)\cdot\left(\frac{1}{c}\frac{2^{n+2}}{(4k+1)\pi}\right)\cdot\left(\frac{1}{c}\frac{2^{n+2}}{(4k+3)\pi}\right)\\ &= \frac{6}{5}\cdot\frac{8}{\pi(4k+1)(4k+3)}.\end{aligned}$$

It follows that for any k, n, we have

$$|a_k(n)| \le \frac{6}{5}\cdot\frac{8}{\pi(4k+1)(4k+3)} =: M_k.$$

Since the sum $\sum_{k=0}^\infty M_k$ converges, we have verified the hypotheses of Tannery's theorem. Hence, taking $n \to \infty$ in (5.26), we get

$$\begin{aligned}1 = \lim_{n\to\infty} \sum_{k=0}^{2^{n-1}-1} a_k(n) &= \sum_{k=0}^\infty \lim_{n\to\infty} a_k(n)\\ &= \sum_{k=0}^\infty \frac{8}{\pi(4k+1)(4k+3)} \quad \Longrightarrow \quad \frac{\pi}{4} = \sum_{k=0}^\infty \frac{2}{(4k+1)(4k+3)}.\end{aligned}$$

The last series is equivalent to Gregory-Leibniz-Madhava's formula because if we use partial fractions, we see that

$$\frac{2}{(4k+1)(4k+3)} = \frac{1}{4k+1} - \frac{1}{4k+3},$$

so when we write out the series term by term we obtain

$$\frac{\pi}{4} = \sum_{k=0}^\infty \left(\frac{1}{4k+1} - \frac{1}{4k+3}\right) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots,$$

which is exactly Gregory-Leibniz-Madhava's formula.

Exercises 5.2.

1. Prove the formula (5.19) by induction on n.
2. Determine the following limit:

$$\lim_{n\to\infty} \left\{ \frac{1}{n^3 \sin\left(\frac{1\cdot 2}{n^3}\right)} + \frac{1}{n^3 \sin\left(\frac{2\cdot 3}{n^3}\right)} + \cdots + \frac{1}{n^3 \sin\left(\frac{n\cdot(n+1)}{n^3}\right)} \right\}.$$

3. (**Partial fraction expansion of $1/\sin^2 x$, Proof I**) Here's Hofbauer's [**102**] derivation of a partial fraction expansion of $1/\sin^2 x$.
   (i) Prove that

$$\frac{1}{\sin^2 x} = \frac{1}{2^{2n}} \sum_{k=0}^{2^n-1} \frac{1}{\sin^2 \frac{x+\pi k}{2^n}}.$$

   (ii) Show that

$$\frac{1}{\sin^2 x} = \frac{1}{2^{2n}} \sum_{k=-2^{n-1}}^{2^{n-1}-1} \frac{1}{\sin^2 \frac{x+\pi k}{2^n}}. \tag{5.28}$$

   (iii) Using Lemma 5.5 prove that $\frac{1}{\sin^2 x} = \lim_{n\to\infty} \sum_{k=-n}^n \frac{1}{(x+\pi k)^2}$. We usually write this as

$$\boxed{\frac{1}{\sin^2 x} = \sum_{k\in\mathbb{Z}} \frac{1}{(x+\pi k)^2}.} \tag{5.29}$$

4. (**Partial fraction expansion of $1/\sin^2 x$, Proof II**) Give another proof of (5.29) using Tannery's theorem and the formula (5.28).
5. (**Euler's sum for $\pi^2/6$, Proof IV**) In this problem we derive Euler's sum via an old argument found in Thomas John l'Anson Bromwich's (1875–1929) book [**41**, p. 218–19] (cf. similar ideas found in [**6**], [**179**], [**123**], [**249**, Problem 145]).
   (i) Recall from Problem 4 in Exercises 4.7 that for any $n \in \mathbb{N}$ and $x \in \mathbb{R}$,

$$\sin nx = \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} (-1)^k \binom{n}{2k+1} \cos^{n-2k-1} x \, \sin^{2k+1} x.$$

Using this formula, prove that if $\sin x \neq 0$, then

$$\sin(2n+1)x = \sin^{2n+1} x \sum_{k=0}^{n} (-1)^k \binom{2n+1}{2k+1} (\cot^2 x)^{n-k}.$$

(ii) Prove that if $n \in \mathbb{N}$, then the roots of $\sum_{k=0}^{n} (-1)^k \binom{2n+1}{2k+1} t^{n-k} = 0$ are the n numbers $t = \cot^2 \frac{m\pi}{2n+1}$ where $m = 1, 2, \ldots, n$.

(iii) Prove that if $n \in \mathbb{N}$, then

$$\sum_{k=1}^{n} \cot^2 \frac{k\pi}{2n+1} = \frac{n(2n-1)}{3}. \tag{5.30}$$

Suggestion: Recall that if $p(t)$ is a polynomial of degree n with roots $r_1, \ldots, r_n$, then $p(t) = a(t-r_1)(t-r_2)\cdots(t-r_n)$ for a constant a. What's the coefficient of t^1 if you multiply out $a(t-r_1)\cdots(t-r_n)$?

(iv) From the identity (5.30), derive Euler's sum.

6. (**Euler's sum for $\pi^2/6$, Proof V**) Here's another proof (cf. [**102**])!

(i) Use (5.29) to prove that for any $n \in \mathbb{N}$,

$$\frac{1}{\sin^2 x} = \frac{1}{n^2} \sum_{m=0}^{n-1} \frac{1}{\sin^2 \frac{x+\pi m}{n}}. \tag{5.31}$$

Suggestion: Replace x with $\frac{x+\pi m}{n}$ in (5.29) and sum from $m = 0$ to $n-1$.

(ii) Take the $m = 0$ term in (5.31) to the left, replace n by $2n+1$, and then take $x \to 0$ to derive the identity

$$\sum_{k=1}^{n} \frac{1}{\sin^2 \frac{\pi k}{2n+1}} = \frac{2n(n+1)}{3}. \tag{5.32}$$

(iii) From the identity (5.32), derive Euler's sum.

7. In this problem we prove that

$$\zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

(i) Prove that for any nonnegative real numbers $a_1, \ldots, a_n$, we have

$$1 - \sum_{k=1}^{n} a_k + \sum_{1 \leq i < j \leq n} a_i\, a_j - \sum_{1 \leq i < j < k \leq n} a_i a_j a_k \leq \prod_{k=1}^{n} (1 - a_k) \leq 1 - \sum_{k=1}^{n} a_k + \sum_{1 \leq i < j \leq n} a_i\, a_j.$$

(ii) Applying the inequalities in (i) to $\prod_{k=1}^{n}\left(1 - \frac{x^2}{k^2\pi^2}\right)$, prove that $\zeta(4) = \pi^4/90$.

5.3. ★ Euler's formula for $\zeta(2k)$

In Euler's famous 1735 paper *De summis serierum reciprocarum* (On the sums of series of reciprocals), he found not only $\zeta(2)$ but he also explicitly determined $\zeta(n)$ for n even up to $n = 12$, although it is clear from his method that he could get the value of $\zeta(n)$ for any even n. Following G.T. Williams [**247**], we derive Euler's formula for $\zeta(n)$, for all $n \in \mathbb{N}$ even, as a rational multiple of π^n (this proof is by far the most elementary proof I know of).

5.3.1. Williams' formula. In order to find Euler's formula, we need the following theorem, whose proof is admittedly long but it is completely elementary in the sense that it basically uses only high school arithmetic and the most basic facts about series!

THEOREM 5.8 (**Williams' formula**). *For any $k \in \mathbb{N}$ with $k \geq 2$, we have*

$$\boxed{\left(k + \frac{1}{2}\right)\zeta(2k) = \sum_{\ell=1}^{k-1} \zeta(2\ell)\,\zeta(2k-2\ell).}$$

PROOF. Fix $k \in \mathbb{N}$ with $k \geq 2$. Then for $N \in \mathbb{N}$, define

$$a_N := \sum_{\ell=1}^{k-1}\left(\sum_{m=1}^{N}\frac{1}{m^{2\ell}}\right)\left(\sum_{n=1}^{N}\frac{1}{n^{2k-2\ell}}\right) = \sum_{\ell=1}^{k-1}\sum_{m,n=1}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}},$$

where for simplicity in notation, we write the double summation $\sum_{m=1}^{N}\sum_{n=1}^{N}$ as as single entity $\sum_{m,n=1}^{N}$. Since $\zeta(z) = \sum_{n=1}^{\infty} 1/n^z = \lim_{N\to\infty}\sum_{n=1}^{N} 1/n^z$, we have

$$\sum_{\ell=1}^{k-1}\zeta(2\ell)\,\zeta(2k-2\ell) = \lim_{N\to\infty} a_N,$$

so we just have to work out a nice formula for a_N and show that $\lim_{N\to\infty} a_N = \left(k+\frac{1}{2}\right)\zeta(2k)$. To accomplish this we proceed in three steps.

Step 1: We first break up the double sum $\sum_{m,n=1}^{N}$ into two sums, one with $m = n$ and the other with $m \neq n$:

$$a_N = \sum_{\ell=1}^{k-1}\sum_{m=n=1}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}} + \sum_{\ell=1}^{k-1}\sum_{m\neq n}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}},$$

where we denote the summation $\sum_{m,n=1}^{N}$ with identical m, n omitted by $\sum_{m\neq n}^{N}$. When $m = n$, we have

$$\begin{aligned}(5.33)\qquad \sum_{\ell=1}^{k-1}\sum_{m=n=1}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}} &= \sum_{\ell=1}^{k-1}\sum_{n=1}^{N}\frac{1}{n^{2\ell}\,n^{2k-2\ell}}\\ &= \sum_{\ell=1}^{k-1}\sum_{n=1}^{N}\frac{1}{n^{2k}} = (k-1)\sum_{n=1}^{N}\frac{1}{n^{2k}},\end{aligned}$$

and for $m \neq n$, we have

$$\sum_{\ell=1}^{k-1}\sum_{m\neq n}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}} = \sum_{\ell=1}^{k-1}\sum_{m\neq n}^{N}\frac{1}{n^{2k}}\left(\frac{n}{m}\right)^{2\ell} = \sum_{m\neq n}^{N}\frac{1}{n^{2k}}\sum_{\ell=1}^{k-1}\left(\frac{n}{m}\right)^{2\ell}.$$

Recalling the formula for a geometric sum: $\sum_{\ell=1}^{k-1} r^\ell = (r - r^k)/(1-r)$ where $r \neq 1$, we can write

$$\frac{1}{n^{2k}}\sum_{\ell=1}^{k-1}\left(\frac{n}{m}\right)^{2\ell} = \frac{1}{n^{2k}}\frac{(n/m)^2 - (n/m)^{2k}}{1-(n/m)^2} = \frac{n^{2-2k} - m^{2-2k}}{m^2 - n^2}.$$

Therefore,

$$\begin{aligned}(5.34)\qquad \sum_{\ell=1}^{k-1}\sum_{m\neq n}^{N}\frac{1}{m^{2\ell}\,n^{2k-2\ell}} &= \sum_{m\neq n}^{N}\frac{n^{2-2k} - m^{2-2k}}{m^2 - n^2}\\ &= \sum_{m\neq n}^{N}\frac{n^{2-2k}}{m^2-n^2} - \sum_{m\neq n}^{N}\frac{m^{2-2k}}{m^2-n^2}.\end{aligned}$$

The second term is actually the same as the first because

$$-\sum_{m\neq n}^{N} \frac{m^{2-2k}}{m^2-n^2} = \sum_{m\neq n}^{N} \frac{m^{2-2k}}{n^2-m^2} = \sum_{n\neq m}^{N} \frac{n^{2-2k}}{m^2-n^2},$$

where in the last equality we switched the letters m and n. Combining (5.33) with (5.34) we get

$$a_N = (k-1)\sum_{n=1}^{N} \frac{1}{n^{2k}} + 2\sum_{m\neq n}^{N} \frac{n^{2-2k}}{m^2-n^2}. \tag{5.35}$$

Step 2: We now find a nice expression for $2\sum_{m\neq n}^{N} \frac{n^{2-2k}}{m^2-n^2}$. To this end we write this as

$$\begin{aligned}
2\sum_{m\neq n}^{N} \frac{n^{2-2k}}{m^2-n^2} &= 2\sum_{n=1}^{N}\left(\sum_{m=1}^{n-1} + \sum_{m=n+1}^{N}\right)\frac{n^{2-2k}}{m^2-n^2}\\
&= \sum_{n=1}^{N}\frac{1}{n^{2k-1}}\left(\sum_{m=1}^{n-1} + \sum_{m=n+1}^{N}\right)\frac{2n}{m^2-n^2}\\
&= \sum_{n=1}^{N}\frac{1}{n^{2k-1}}\left(\sum_{m=1}^{n-1} + \sum_{m=n+1}^{N}\right)\left(\frac{1}{m-n} - \frac{1}{m+n}\right). \qquad (5.36)
\end{aligned}$$

where at the last step we used partial fractions $\frac{2n}{m^2-n^2} = \frac{1}{m-n} - \frac{1}{m+n}$. We now work out the second sums in (5.36). First observe that

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

Second, observe that

$$\sum_{m=n+1}^{N}\frac{1}{m-n} = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{N-n} = \sum_{m=1}^{N-n}\frac{1}{m}.$$

Third, observe that

$$\begin{aligned}
-\left(\sum_{m=1}^{n-1} + \sum_{m=n+1}^{N}\right)\frac{1}{m+n} &= \frac{1}{n+n} - \sum_{m=1}^{N}\frac{1}{m+n}\\
&= \frac{1}{n+n} - \left(\frac{1}{1+n} + \frac{1}{2+n} + \cdots + \frac{1}{N+n}\right)\\
&= \frac{1}{2n} - \sum_{m=n+1}^{N+n}\frac{1}{m}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left(\sum_{m=1}^{n-1} + \sum_{m=n+1}^{N}\right)&\left(\frac{1}{m-n} - \frac{1}{m+n}\right)\\
&= \left(\frac{1}{n} - \sum_{m=1}^{n}\frac{1}{m}\right) + \sum_{m=1}^{N-n}\frac{1}{m} + \left(\frac{1}{2n} - \sum_{m=n+1}^{N+n}\frac{1}{m}\right)\\
&= \frac{3}{2n} - \sum_{m=1}^{N+n}\frac{1}{m} + \sum_{m=1}^{N-n}\frac{1}{m} \qquad (5.37)\\
&= \frac{3}{2n} + \sum_{m=N-n+1}^{N+n}\frac{1}{m}.
\end{aligned}$$

Thus, by (5.36), we have

$$\begin{aligned}
2\sum_{m\neq n}^{N}\frac{n^{2-2k}}{m^2-n^2} &= \sum_{n=1}^{N}\frac{1}{n^{2k-1}}\left(\frac{3}{2n} + \sum_{m=N-n+1}^{N-n}\frac{1}{m}\right)\\
&= \frac{3}{2}\sum_{n=1}^{N}\frac{1}{n^{2k}} + \sum_{n=1}^{N}\left(\frac{1}{n^{2k-1}}\sum_{m=N-n+1}^{N+n}\frac{1}{m}\right).
\end{aligned}$$

Plugging this into the formula (5.35) for a_N, we obtain

$$a_N = \left(k+\frac{1}{2}\right)\sum_{n=1}^{N}\frac{1}{n^{2k}} + \sum_{n=1}^{N}\left(\frac{1}{n^{2k-1}}\sum_{m=N-n+1}^{N+n}\frac{1}{m}\right).$$

Therefore, $\lim a_N = (k-1/2)\zeta(2k)$ provided we can show that

$$0 = \lim_{N\to\infty}\sum_{n=1}^{N}\left(\frac{1}{n^{2k-1}}\sum_{m=N-n+1}^{N+n}\frac{1}{m}\right). \qquad (5.38)$$

Step 3: We prove the limit (5.38) (see Problem 1 for a proof using Tannery's theorem). To this end, observe that

$$\begin{aligned}
\sum_{m=N-n+1}^{N+n}\frac{1}{m} &= \frac{1}{N-n+1} + \frac{1}{N-n+2} + \cdots + \frac{1}{N+n}\\
&\leq \frac{1}{N-n+1} + \frac{1}{N-n+1} + \cdots + \frac{1}{N-n+1} = \frac{2n}{N-n+1}.
\end{aligned}$$

Therefore,

$$\sum_{n=1}^{N}\left(\frac{1}{n^{2k-1}}\sum_{m=N-n+1}^{N+n}\frac{1}{m}\right) \leq 2\sum_{n=1}^{N}\left(\frac{n}{n^{2k-1}}\frac{1}{N-n+1}\right) = 2\sum_{n=1}^{N}\frac{1}{n^2(N-n+1)},$$

where recall that $k \geq 2$ so that $\frac{n}{n^{2k-1}} \leq \frac{1}{n^2}$. Using partial fractions we see that

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

Thus,

$$\begin{aligned}\frac{1}{n^2(N-n+1)} &= \frac{1}{n}\cdot\frac{1}{n(N-n+1)} = \frac{1}{n}\cdot\frac{1}{N+1}\left(\frac{1}{n}+\frac{1}{N-n+1}\right)\\ &= \frac{1}{N+1}\left(\frac{1}{n^2}+\frac{1}{n(N-n+1)}\right)\\ &= \frac{1}{N+1}\left(\frac{1}{n^2}+\frac{1}{N+1}\left(\frac{1}{n}+\frac{1}{N-n+1}\right)\right)\\ &= \frac{1}{N+1}\cdot\frac{1}{n^2}+\frac{1}{(N+1)^2}(1+1)\\ &\le \frac{1}{N+1}\cdot\frac{1}{n^2}+\frac{2}{(N+1)^2}.\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{n=1}^{N}\left(\frac{1}{n^{2k-1}}\sum_{m=N-n+1}^{N+n}\frac{1}{m}\right) &= 2\sum_{n=1}^{N}\frac{1}{n^2(N-n+1)}\\ &\le \frac{2}{N+1}\sum_{n=1}^{N}\frac{1}{n^2}+\sum_{n=1}^{N}\frac{2}{(N+1)^2} \le \frac{2\pi^2/6}{N+1}+\frac{2N}{(N+1)^2}.\end{aligned}$$

Taking $N\to\infty$ proves (5.38) and completes our proof. □

In particular, setting $k=2$ we see that $\frac{5}{2}\zeta(4)=\zeta(2)^2$. Thus, $\zeta(4)=\frac{2}{5}\frac{\pi^4}{36}=\frac{\pi^2}{90}$. Taking $k=3$, we get

$$\frac{7}{2}\zeta(6)=\zeta(2)\zeta(4)+\zeta(4)\zeta(2)=2\zeta(2)\zeta(4)=2\cdot\frac{\pi^2}{6}\cdot\frac{\pi^4}{90},$$

which after doing the algebra, we get $\zeta(6)=\pi^2/945$. Thus,

$$\boxed{\frac{\pi^4}{90}=\sum_{n=1}^{\infty}\frac{1}{n^4}\quad,\qquad\frac{\pi^6}{945}=\sum_{n=1}^{\infty}\frac{1}{n^6}.}$$

We can also derive explicit formulas for $\zeta(2k)$ for all $k\in\mathbb{N}$.

5.3.2. Euler's formula for $\zeta(2k)$. To this end, we first define a sequence $C_1, C_2, C_3, \ldots$ by $C_1=\frac{1}{12}$, and for $k\ge 2$, we define

$$C_k=-\frac{1}{2k+1}\sum_{\ell=1}^{k-1}C_\ell C_{k-\ell}. \tag{5.39}$$

The first few C_k's are

$$C_1=\frac{1}{12},\quad C_2=-\frac{1}{720},\quad C_3=\frac{1}{30240},\quad C_4=-\frac{1}{1209600},\ldots.$$

The numbers C_k are rational numbers (easily proved by induction) and are related to the **Bernoulli numbers** to be covered in Section 6.8, but it's not necessary to know this.[6] We are now ready to prove ...

[6] Explicitly, $C_k=B_{2k}/(2k)!$ but this formula is not needed.

Theorem 5.9 (**Euler's formulæ**). *For any $k \in \mathbb{N}$, we have*

$$\boxed{\sum_{n=1}^{\infty} \frac{1}{n^{2k}} = (-1)^{k-1} \frac{(2\pi)^{2k}\, C_k}{2} \quad ; \quad \textit{that is,} \;\; \zeta(2k) = (-1)^{k-1} \frac{(2\pi)^{2k}\, C_k}{2}.} \tag{5.40}$$

Proof. When $k = 1$, we have

$$(-1)^{k-1} \frac{(2\pi)^{2k}\, C_k}{2} = \frac{(2\pi)^2\,(1/12)}{2} = \frac{\pi^2}{6} = \zeta(2),$$

so our theorem holds when $k = 1$. Let $k \geq 2$ and assume our theorem holds for all natural numbers less than k; we shall prove it holds for k. Using Williams' formula and the induction hypothesis, we see that

$$\begin{aligned}
\left(k + \frac{1}{2}\right) \zeta(2k) &= \sum_{\ell=1}^{k-1} \zeta(2\ell)\zeta(2k - 2\ell) \\
&= \sum_{\ell=1}^{k-1} \left((-1)^{\ell-1} \frac{(2\pi)^{2\ell}\, C_\ell}{2} \right) \left((-1)^{k-\ell-1} \frac{(2\pi)^{2k-2\ell}\, C_{k-2}}{2} \right) \\
&= \sum_{\ell=1}^{k-1} \left((-1)^{k-2} \frac{(2\pi)^{2k}\, C_\ell C_{k-\ell}}{4} \right) \\
&= (-1)^{k-2} \frac{(2\pi)^{2k}}{4} \sum_{\ell=1}^{k-1} C_\ell C_{k-\ell} \\
&= (-1)^{k-1} \frac{(2\pi)^{2k}}{4} (2k + 1) C_k.
\end{aligned}$$

Dividing everything by $(k + 1/2) = (1/2)(2k + 1)$ and using the formula (5.39) for C_k proves our result for k. □

As a side note, we remark that (5.40) shows that $\zeta(2k)$ is a rational number times π^{2k}; in particular, since π is transcendental (see, for example, [**162, 163, 136**]) it follows that $\zeta(n)$ is transcendental for n even. One may ask if there are similar expressions like (5.40) for sums of the reciprocals of the *odd* powers (e.g. $\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3}$). Unfortunately, there are no known formulas! Moreover, it is not even known if $\zeta(n)$ is transcendental for n odd and in fact, of all odd numbers only $\zeta(3)$ is known without a doubt to be irrational; this was proven by Roger Apéry (1916–1994) in 1979 (see [**29**], [**230**])!

Exercises 5.3.

1. Prove (5.38) using Tannery's theorem.
2. (Cf. [**116**]) Let $H_n = \sum_{m=1}^{n} \frac{1}{m}$, the n-th partial sum of the harmonic series. In this problem we prove the equalities:

$$\boxed{\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} = \sum_{n=1}^{\infty} \frac{H_n}{(n+1)^2} = \frac{1}{2} \sum_{n=1}^{\infty} \frac{H_n}{n^2}.} \tag{5.41}$$

   (a) Prove that for $N \in \mathbb{N}$,

$$\sum_{m,n=1}^{N} \frac{1}{mn(m+n)} = \sum_{m=1}^{N} \frac{H_m}{m^2} = \sum_{m=1}^{N} \frac{1}{m^3} + \sum_{n=1}^{N-1} \frac{H_k}{(k+1)^2},$$

where the notation $\sum_{m,n=1}^N$ is as in the proof of Williams' theorem. Suggestion: For the first equality, use that $\frac{1}{mn(m+n)} = \frac{1}{m^2}\left(\frac{1}{n} - \frac{1}{m+n}\right)$.

(b) Now prove that for $N \in \mathbb{N}$,

$$\sum_{m,n=1}^{N} \frac{1}{mn(m+n)} = 2\sum_{m=1}^{N}\sum_{n=1}^{N} \frac{1}{m(m+n)^2}.$$

Suggestion: Use that $\frac{1}{mn(m+n)} = \frac{1}{m(m+n)^2} + \frac{1}{n(m+n)^2}$.

(c) In Part (b), instead of using n as the inner summation variable on the right-hand side, change to $k = m+n-1$ and in doing so, prove that

$$\sum_{m,n=1}^{N} \frac{1}{mn(m+n)} = 2\sum_{m=1}^{N}\sum_{k=m}^{N} \frac{1}{m(k+1)^2} + b_N, \quad \text{where } b_N = 2\sum_{m=1}^{N}\sum_{k=N+1}^{m+N-1} \frac{1}{m(k+1)^2}.$$

(d) Show that $\sum_{m=1}^{N}\sum_{k=m}^{N} \frac{1}{m(k+1)^2} = \sum_{k=1}^{N} \frac{H_k}{(k+1)^2}$ and that $b_N \to 0$ as $N \to \infty$. Now prove (5.41).

3. (**Euler's sum for $\pi^2/6$, Proof VI**) In this problem we prove Euler's formula for $\pi^2/6$ by *carefully* squaring Gregory-Leibniz-Madhava's formula for $\pi/4$; thus, taking Gregory-Leibniz-Madhava's formula as "given," we derive Euler's formula.[7] The proof is very much in the same spirit as the proof of Williams' formula; see Section 6.11 for another, more systematic, proof.

(i) Given $N \in \mathbb{N}$, prove that

$$\left(\sum_{m=0}^{N} \frac{(-1)^m}{(2m+1)}\right)\left(\sum_{n=0}^{N} \frac{(-1)^n}{(2n+1)}\right) = \sum_{n=0}^{N} \frac{1}{(2n+1)^2} + \sum_{m\neq n}^{N} \frac{(-1)^{m+n}}{(2m+1)(2n+1)}$$

where the notation $\sum_{m\neq n}^{N}$ is as in the proof of Williams' theorem.

(ii) For $m \neq n$, prove that[8]

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

then use this identity to prove that

$$\begin{aligned}\sum_{m\neq n}^{N} \frac{(-1)^{m+n}}{(2m+1)(2n+1)} &= 2\sum_{m\neq n}^{N} \frac{(-1)^{m+n}}{2n+1}\cdot\frac{2m+1}{(2m+1)^2-(2n+1)^2}\\ &= 2\sum_{n=0}^{N}\frac{(-1)^n}{2n+1}\left(\sum_{m=0}^{n-1} + \sum_{m=n+1}^{N}\right)(-1)^m\frac{2m+1}{(2m+1)^2-(2n+1)^2}.\end{aligned}$$

(iii) Prove that

$$4\left(\sum_{m=0}^{n-1} + \sum_{m=n+1}^{N}\right)(-1)^m\frac{2m+1}{(2m+1)^2-(2n+1)^2} = -\frac{(-1)^n}{2n+1} + (-1)^N\sum_{m=N-n+1}^{N+n+1}\frac{1}{m}.$$

Suggestion: Note that $4\frac{2m+1}{(2m+1)^2-(2n+1)^2} = \frac{2m+1}{(m+n+1)(m-n)} = \frac{1}{m-n} + \frac{1}{m+n+1}$.

[7] Actually, this works in reverse: We can just as well take Euler's formula as "given," and then derive Gregory-Leibniz-Madhava's formula!

[8] Alternatively, one can prove that $\frac{1}{(2m+1)(2n+1)} = \frac{1}{2(m-n)(2n+1)} + \frac{1}{2(n-m)(2m+1)}$ and use this decomposition to simplify $\sum_{m\neq n}^{N} \frac{(-1)^{m+n}}{(2m+1)(2n+1)}$. However, if you do Problem 4 to follow, in your proof you will run into the decomposition appearing in Part (b) above!

(iv) Prove that

$$\left(\sum_{m=0}^{N}\frac{(-1)^m}{(2m+1)}\right)\left(\sum_{n=0}^{N}\frac{(-1)^n}{(2n+1)}\right)=b_N+\frac{1}{2}\sum_{n=0}^{N}\frac{1}{(2n+1)^2},$$

$$\text{where } b_N=\frac{1}{2}\sum_{n=0}^{N}\frac{(-1)^{N+n}}{(2n+1)}\left(\sum_{m=N-n+1}^{N+n+1}\frac{1}{m}\right).$$

(v) Prove that $b_N\to 0$ as $N\to\infty$, and conclude that $(\pi/4)^2=\frac{1}{2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}$. Finally, derive Euler's formula for $\pi^2/6$.

4. (**Williams' other formula**) For any $k\in\mathbb{N}$, define

$$\xi(k)=\sum_{n=0}^{\infty}(-1)^n\frac{1}{(2n+1)^k}.$$

For example, by Gregory-Leibniz-Madhava's formula we know that $\xi(1)=\pi/4$. Prove that for any $k\in\mathbb{N}$ with $k\ge 2$, we have

$$\left(k-\frac{1}{2}\right)\sum_{n=0}^{\infty}\frac{1}{(2n+1)^{2k}}=\sum_{\ell=0}^{k-1}\xi(2\ell+1)\,\xi(2k-2\ell-1).$$

This formula also holds when $k=1$, for it's simply the identity $\frac{1}{2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}=\xi(1)^2$, a fact established in Problem 3. Suggestion: Imitate the proof of Williams' formula. You will see that ideas from Problem 3 will also be useful.

5. (Cf. [**247**], [**37**], [**117**]) Let $H_n=\sum_{m=1}^{n}\frac{1}{m}$, the n-th partial sum of the harmonic series. In this problem we prove that for any $k\in\mathbb{N}$ with $k\ge 2$, we have

$$\boxed{(k+2)\,\zeta(k+1)=\sum_{\ell=1}^{k-2}\zeta(k-\ell)\,\zeta(\ell+1)+2\sum_{n=1}^{\infty}\frac{H_n}{n^k},} \tag{5.42}$$

a formula due to Euler (no surprise!). The proof is very similar to the proof of Williams' formula, with some twists of course. You may proceed as follows.

(i) For $N\in\mathbb{N}$, define

$$a_N=\sum_{\ell=1}^{k-2}\left(\sum_{m=1}^{N}\frac{1}{m^{k-\ell}}\right)\left(\sum_{n=1}^{N}\frac{1}{n^{\ell+1}}\right)=\sum_{m,n=1}^{N}\sum_{\ell=1}^{k-2}\frac{1}{m^{k-\ell}\,n^{\ell+1}},$$

Summing the geometric sum $\sum_{\ell=1}^{k-2}\frac{1}{m^{k-\ell}\,n^{\ell+1}}=\frac{1}{m^k\,n}\sum_{\ell=1}^{k-2}(m/n)^{\ell}$, prove that

$$a_N=(k-2)\sum_{n=1}^{N}\frac{1}{n^{k+1}}+2\sum_{m\neq n}^{N}\frac{1}{n^{k-1}\,m\,(m-n)},$$

where the notation $\sum_{m\neq n}^{N}$ is as in the proof of Williams' theorem.

(ii) Prove that

$$\sum_{m\neq n}^{N}\frac{1}{n^{k-1}\,m\,(m-n)}=\sum_{n=1}^{N}\frac{1}{n^k}\left(\sum_{m=1}^{n-1}+\sum_{m=n+1}^{N}\right)\left(\frac{1}{m-n}-\frac{1}{m}\right)$$

(iii) Prove that

$$\left(\sum_{m=1}^{n-1}+\sum_{m=n+1}^{N}\right)\left(\frac{1}{m-n}-\frac{1}{m}\right)=\frac{2}{n}-H_n-\sum_{m=N-n+1}^{N}\frac{1}{m}.$$

Suggestion: The computation around (5.37) might be helpful.

(iv) Prove that

$$a_N = (k+2)\sum_{n=1}^{N} \frac{1}{n^{k+1}} - 2\sum_{n=1}^{N} \frac{H_n}{n^k} - b_N, \quad \text{where} \quad b_N = 2\sum_{n=1}^{N} \frac{1}{n^k}\left(\sum_{m=N-n+1}^{N} \frac{1}{m}\right).$$

(v) Prove that $b_N \to 0$ as $N \to \infty$, and conclude that (5.42) holds.

6. (Cf. [**116**], [**117**]) Here's are a couple applications of (5.42). First, use (5.42) to give a quick proof of (5.41). Second, prove that

$$\boxed{\frac{\pi^4}{72} = \sum_{n=1}^{\infty} \frac{1}{n^3}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right).}$$

Part 2

Extracurricular activities

CHAPTER 6

Advanced theory of infinite series

Ut non-fınitam Seriem fınita cöercet,
Summula, & in nullo limite limes adest:
Sic modico immensi vestigia Numinis haerent
Corpore, & angusto limite limes abest.
Cernere in immenso parvum, dic, quanta voluptas!
In parvo immensum cernere, quanta, Deum.

Even as the finite encloses an infinite series
And in the unlimited limits appear,
So the soul of immensity dwells in minutia
And in the narrowest limits no limit in here.
What joy to discern the minute in infinity!
The vast to perceive in the small, what divinity!
Jacob Bernoulli (1654-1705) Ars Conjectandi.[**216**, p. 271]

This chapter is about going in-depth into the theory and application of infinite series. One infinite series that will come up again and again in this chapter and the next chapter as well, is the Riemann zeta function

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z},$$

introduced in Section 4.6. Amongst many other things, in this chapter we'll see how to write some well-known constants in terms of the Riemann zeta function; e.g. we'll derive the following neat formula for our friend $\log 2$ (§ 6.5):

$$\boxed{\log 2 = \sum_{n=2}^{\infty} \frac{1}{2^n}\, \zeta(n),}$$

another formula for our friend the Euler-Mascheroni constant (§ 6.9):

$$\boxed{\gamma = \sum_{n=2}^{\infty} \frac{(-1)^n}{n} \zeta(n),}$$

and two more formulas involving our most delicious friend π (see §'s 6.10 and 6.11):

$$\boxed{\pi = \sum_{n=2}^{\infty} \frac{3^n - 1}{4^n}\, \zeta(n+1) \quad , \quad \frac{\pi^2}{6} = \zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots .}$$

We'll also re-derive Gregory-Leibniz-Madhava's formula (§ 6.10)

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + - \cdots ,}$$

and Machin's formula which started the "decimal place race" of computing π (§ 6.10):

$$\boxed{\pi = 4\arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right) = 4\sum_{n=0}^\infty \frac{(-1)^n}{(2n+1)}\left(\frac{4}{5^{2n+1}} - \frac{1}{239^{2n+1}}\right).}$$

Leibniz's formula for $\pi/4$ is an example of an "alternating series". We study these types of series in Section 6.1. In Section 6.2 and Section 6.3 we look at the ratio and root tests, which you are probably familiar with from elementary calculus. In Section 6.4 we look at power series and prove some pretty powerful properties of power series. The formulas for $\log 2$, γ, and the formula $\pi = \sum_{n=2}^\infty \frac{3^n-1}{4^n}\zeta(n+1)$ displayed above are proved using a famous theorem called the *Cauchy double series theorem.* This theorem, and double sequences and series in general, are the subject of Section 6.5. In Section 6.6 we investigate rearranging (that is, mixing up the order of the terms in a) series. Here's an interesting question: Does the series

$$\sum_{p \text{ is prime}} \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{23} + \frac{1}{29} + \cdots$$

converge or diverge? For the answer, see Section 6.7. In elementary calculus, you probably never seen the power series representations of tangent and secant. This is because these series are somewhat sophisticated mathematically speaking. In Section 6.8 we shall derive the power series representations

$$\boxed{\tan z = \sum_{n=1}^\infty (-1)^{n-1}\frac{2^{2n}(2^{2n}-1)\,B_{2n}}{(2n)!}\, z^{2n-1},}$$

and

$$\boxed{\sec z = \sum_{n=0}^\infty (-1)^n \frac{E_{2n}}{(2n)!}\, z^{2n}.}$$

Here, the B_{2n}'s are called "Bernoulli numbers" and the E_{2n}'s are called "Euler numbers," which are certain numbers having extraordinary properties. Although you've probably never seen the tangent and secant power series, you might have seen the logarithmic, binomial, and arctangent series:

$$\boxed{\log(1+z) = \sum_{n=1}^\infty \frac{(-1)^{n-1}}{n} z^n\ ,\ (1+z)^\alpha = \sum_{n=0}^\infty \binom{\alpha}{n} z^n\ ,\ \arctan z = \sum_{n=0}^\infty (-1)^n \frac{z^{2n+1}}{2n+1}}$$

where $\alpha \in \mathbb{R}$. You most likely used calculus (derivatives and integrals) to derive these formulæ. In Section 6.9 we shall derive these formulæ without any calculus. Finally, in Sections 6.10 and 6.11 we derive many incredible and awe-inspiring formulæ involving π. In particular, we again look at the Basel problem.

Chapter 6 objectives: The student will be able to ...

- determine the convergence, and radius and interval of convergence, for an infinite series and power series, respectively, using various tests, e.g. Dirichlet, Abel, ratio, root, and others.
- apply Cauchy's double series theorem and know how it relates to rearrangement, and multiplication and composition of power series.
- identify series formulæ for the various elementary functions (binomial, arctangent, etc.) and for π.

6.1. Summation by parts, bounded variation, and alternating series

In elementary calculus, you studied "integration by parts," a formula I'm sure you used quite often trying to integrate tricky integrals. In this section we study a discrete version of the integration by parts formula called "summation by parts," which is used to sum tricky summations! Summation by parts has broad applications, including finding sums of powers of integers and to derive some famous convergence tests for series, the Dirichlet and Abel tests.

6.1.1. Summation by parts and Abel's lemma.

Here is the famous summation by parts formula. The formula is complicated, but the proof is simple.

Theorem 6.1 (**Summation by parts**). *For any complex sequences $\{a_n\}$ and $\{b_n\}$, we have*

$$\sum_{k=m}^{n} b_{k+1}(a_{k+1} - a_k) + \sum_{k=m}^{n} a_k(b_{k+1} - b_k) = a_{n+1}b_{n+1} - a_m b_m.$$

Proof. Combining the two terms on the left, we obtain

$$\sum_{k=m}^{n} \Big[b_{k+1}a_{k+1} - b_{k+1}a_k + a_k b_{k+1} - a_k b_k\Big] = \sum_{k=m}^{n} \big(b_{k+1}a_{k+1} - a_k b_k\big).$$

This is a telescoping sum, and it simplifies to $a_{n+1}b_{n+1} - a_m b_m$ after all the cancellations. □

As a corollary, we get Abel's lemma named after Niels Abel[1] (1802–1829).

Corollary 6.2 (**Abel's lemma**). *Let $\{a_n\}$ and $\{b_n\}$ be any complex sequences and let s_n denote the n-th partial sum of the series corresponding to the sequence $\{a_n\}$. Then for any $m < n$ we have*

$$\sum_{k=m+1}^{n} a_k b_k = s_n b_n - s_m b_m - \sum_{k=m}^{n-1} s_k(b_{k+1} - b_k).$$

Proof. Applying the summation by parts formula to the sequences $\{s_n\}$ and $\{b_n\}$, we obtain

$$\sum_{k=m}^{n-1} b_{k+1}(s_{k+1} - s_k) + \sum_{k=m}^{n-1} s_k(b_{k+1} - b_k) = s_n b_n - s_m b_m.$$

Since $a_{k+1} = s_{k+1} - s_k$, we conclude that

$$\sum_{k=m}^{n-1} b_{k+1}a_{k+1} + \sum_{k=m}^{n-1} s_k(b_{k+1} - b_k) = s_n b_n - s_m b_m.$$

Replacing k with $k - 1$ in the first sum and bringing the second sum to the right, we get our result. □

Summation by parts is a very useful tool. We shall apply it to find sums of powers of integers (cf. [**254**], [**77**]); see the exercises for more applications.

[1] *Abel has left mathematicians enough to keep them busy for 500 years. Charles Hermite (1822–1901), in "Calculus Gems"* [**210**].

6.1.2. Sums of powers of integers.

Example 6.1. Let $a_k = k$ and $b_k = k$. Then each of the differences $a_{k+1} - a_k$ and $b_{k+1} - b_k$ equals 1, so by summation by parts, we have

$$\sum_{k=1}^{n}(k+1) + \sum_{k=1}^{n} k = (n+1)(n+1) - 1 \cdot 1.$$

This sum reduces to

$$2\sum_{k=1}^{n} k = (n+1)^2 - n - 1 = n(n+1),$$

which gives the well-known result:

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

Example 6.2. Now let $a_k = k^2 - k = k(k-1)$ and $b_k = k - 1/2$. In this case, $a_{k+1} - a_k = (k+1)k - k(k-1) = 2k$ and $b_{k+1} - b_k = 1$, so by the summation by parts formula,

$$\sum_{k=1}^{n}\left(k + \frac{1}{2}\right)(2k) + \sum_{k=1}^{n}(k^2 - k)(1) = (n+1)n \cdot \left(n + \frac{1}{2}\right).$$

The first sum on the left contains the sum $\sum_{k=1}^{n} k$ and the second one contains the negative of the same sum. Cancelling, we get

$$3\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{2},$$

which gives the well-known result:

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

Example 6.3. For our final result, let $a_k = k^2$ and $b_k = (k-1)^2$. Then $a_{k+1} - a_k = (k+1)^2 - k^2 = 2k+1$ and $b_{k+1} - b_k = 2k - 1$, so by the summation by parts formula,

$$\sum_{k=1}^{n}(k+1)^2(2k+1) + \sum_{k=1}^{n} k^2(2k-1) = (n+1)^2 \cdot n^2.$$

After some work simplifying the left-hand side, we get

$$1^3 + 2^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}.$$

6.1.3. Sequences of bounded variation and Dirichlet's test. A sequence $\{a_n\}$ of complex numbers is said to be of **bounded variation** if

$$\sum_{n=1}^{\infty} |a_{n+1} - a_n| < \infty.$$

Typical examples of a sequences of of bounded variation are bounded monotone sequences of real numbers. A nice property of general sequences of bounded variation is that they always converge. We prove these facts in the following

Proposition 6.3. *Any sequence of bounded variation converges. Moreover, any bounded monotone sequence is of bounded variation.*

Proof. Let $\{a_n\}$ be of bounded variation. Given $m < n$, we can write $a_n - a_m$ as a telescoping sum:

$$\begin{aligned} a_n - a_m = (a_{m+1} - a_m) + (a_{m+2} - a_{m+1}) + \cdots \\ + (a_{n-1} - a_{n-2}) + (a_n - a_{n-1}) = \sum_{k=m}^{n} (a_{k+1} - a_k). \end{aligned}$$

Hence,

$$|a_n - a_m| \leq \sum_{k=m}^{n} |a_{k+1} - a_k|.$$

By assumption, the sum $\sum_{k=1}^{\infty} |a_{k+1} - a_k|$ converges, so the sum on the right-hand side of this inequality can be made arbitrarily small as $m, n \to \infty$ (Cauchy's criterion for series). Thus, $\{a_n\}$ is Cauchy and hence converges.

Now let $\{a_n\}$ be a nondecreasing and bounded sequence. We shall prove that this sequence is of bounded variation; the proof for a nonincreasing sequence is similar. In this case, we have $a_n \leq a_{n+1}$ for each n, so for each n,

$$\begin{aligned} \sum_{k=1}^{n} |a_{k+1} - a_k| = \sum_{k=1}^{n} (a_{k+1} - a_k) = (a_2 - a_1) + (a_3 - a_2) \\ + \cdots + (a_n - a_{n-1}) + (a_{n+1} - a_n) = a_{n+1} - a_1, \end{aligned}$$

since the sum telescoped. The sequence $\{a_n\}$ is by assumption bounded, so it follows that the partial sums of the infinite series $\sum_{n=1}^{\infty} |a_{n+1} - a_n|$ are bounded, hence the series must converge by the nonnegative series test (Theorem 3.20). □

Here's a useful test named after Johann Dirichlet (1805–1859).

Theorem 6.4 (**Dirichlet's test**). *Suppose that the partial sums of the series $\sum a_n$ are uniformly bounded (although the series $\sum a_n$ may not converge). Then for any sequence $\{b_n\}$ that is of bounded variation and converges to zero, the series $\sum a_n b_n$ converges. In particular, the series $\sum a_n b_n$ converges if $\{b_n\}$ is a monotone sequence of real numbers approaching zero.*

Proof. The trick to use Abel's lemma to rewrite $\sum a_n b_n$ in terms of an absolutely convergent series. Define $a_0 = 0$ (so that $s_0 = a_0 = 0$) and $b_0 = 0$. Then setting $m = 0$ in Abel's lemma, we can write

$$\sum_{k=1}^{n} a_k b_k = s_n b_n - \sum_{k=1}^{n-1} s_k (b_{k+1} - b_k). \tag{6.1}$$

Now we are given two facts: The first is that the partial sums $\{s_n\}$ are bounded, say by a constant C, and the second is that the sequence $\{b_n\}$ is of bounded variation and converges to zero. Since $\{s_n\}$ is bounded and $b_n \to 0$ it follows that $s_n b_n \to 0$. Since $|s_n| \leq C$ for all n and $\{b_n\}$ is of bounded variation, the sum $\sum_{k=1}^{\infty} s_k (b_{k+1} - b_k)$ is absolutely convergent:

$$\sum_{k=1}^{\infty} |s_k (b_{k+1} - b_k)| \leq C \sum_{k=1}^{\infty} |b_{k+1} - b_k| < \infty.$$

Therefore, taking $n \to \infty$ in (6.1) it follows that the sum $\sum a_k b_k$ converges (and equals $\sum_{k=1}^\infty s_k(b_{k+1} - b_k)$), and our proof is complete. □

Example 6.4. For each $x \in (0, 2\pi)$, determine the convergence of the series

$$\sum_{n=1}^\infty \frac{e^{inx}}{n}.$$

To do so, we let $a_n = e^{inx}$ and $b_n = 1/n$. Since $\{1/n\}$ is a monotone sequence converging to zero, by Dirichlet's test, if we can prove that the partial sums of $\sum e^{inx}$ are bounded, then $\sum_{n=1}^\infty \frac{e^{inx}}{n}$ converges. To establish this boundedness, we observe that

$$\sum_{n=1}^m e^{inx} = e^{ix} \frac{1 - e^{imx}}{1 - e^{inx}},$$

where we summed $\sum_{n=1}^m (e^{ix})^n$ via the geometric progression (2.3). Hence,

$$\left| \sum_{n=1}^m e^{inx} \right| \leq \left| \frac{1 - e^{imx}}{1 - e^{inx}} \right| \leq \frac{1 + |e^{imx}|}{1 - e^{inx}|} = \frac{2}{|1 - e^{ix}|}.$$

Since $1 - e^{ix} = e^{ix/2}(e^{-ix/2} - e^{ix/2}) = -2ie^{ix/2} \sin(x/2)$, we see that

$$|1 - e^{ix}| = 2|\sin(x/2)| \quad \Longrightarrow \quad \left| \sum_{n=1}^m e^{inx} \right| \leq \frac{1}{\sin(x/2)}.$$

Thus, for each $x \in (0, 2\pi)$, by Dirichlet's test, given any sequence $\{b_n\}$ of bounded variation that converges to zero, the sum $\sum_{n=1}^\infty b_n e^{inx}$ converges. In particular, $\sum_{n=1}^\infty \frac{e^{inx}}{n}$ converges, and more generally, $\sum_{n=1}^\infty \frac{e^{inx}}{n^p}$ converges for any $p > 0$. Taking real and imaginary parts shows that for any $x \in (0, 2\pi)$,

$$\sum_{n=1}^\infty \frac{\cos nx}{n} \quad \text{and} \quad \sum_{n=1}^\infty \frac{\sin nx}{n} \quad \text{converge.}$$

Before going to other tests, it might be interesting to note that we can determine the convergence of the series $\sum_{n=1}^\infty \frac{\cos nx}{n}$ without using the fancy technology of Dirichlet's test. To this end, observe that from the addition formulas for $\sin(n \pm 1/2)x$, we have

$$\cos nx = \frac{\sin(n + 1/2)x - \sin(n - 1/2)x}{2\sin(x/2)},$$

which implies that, after gathering like terms,

$$\begin{aligned}\sum_{n=1}^{m}\frac{\cos nx}{n} &= \frac{1}{2\sin(x/2)}\sum_{n=1}^{m}\frac{\sin(n+1/2)x-\sin(n-1/2)x}{n}\\ &= \frac{1}{2\sin(x/2)}\Big(\frac{\sin(3x/2)-\sin(x/2)}{1}+\frac{\sin(5x/2)-\sin(3x/2)}{3}\\ &\qquad\qquad +\cdots+\frac{\sin(m+1/2)x-\sin(m-1/2)x}{m}\Big)\\ &= \frac{1}{2\sin(x/2)}\Big(-\sin(x/2)+\frac{\sin(m+1/2)x}{m}+\sum_{n=1}^{m-1}\sin(n+1/2)x\Big(\frac{1}{n}-\frac{1}{n+1}\Big)\Big)\\ &= \frac{1}{2\sin(x/2)}\Big(-\sin(x/2)+\frac{\sin(m+1/2)x}{m}+\sum_{n=1}^{m-1}\sin(n+1/2)x\frac{1}{n(n+1)}\Big).\end{aligned}$$

Therefore,

$$\frac{1}{2}+\sum_{n=1}^{m}\frac{\cos nx}{n}=\frac{\sin(m+1/2)x}{2m\sin(x/2)}+\sum_{n=1}^{m-1}\Big(\frac{\sin(n+1/2)x}{2\sin(x/2)}\cdot\frac{1}{n(n+1)}\Big). \tag{6.2}$$

Since the sine is always bounded by 1 and $\sum 1/n(n+1)$ converges, it follows that as $m\to\infty$, the first term on the right of (6.2) tends to zero while the summation on the right of (6.2) converges; in particular, the series in question converges, and we get the following pretty formula:

$$\frac{1}{2}+\sum_{n=1}^{\infty}\frac{\cos nx}{n}=\frac{1}{2\sin(x/2)}\sum_{n=1}^{\infty}\frac{\sin(n+1/2)x}{n(n+1)}\quad,\quad x\in(0,2\pi).$$

In Example 6.39 of Section 6.9, we'll show that $\sum_{n=1}^{\infty}\frac{\cos nx}{n}=\log(2\sin(x/2))$.

6.1.4. Alternating series tests, $\log 2$, and the irrationality of e.

As a direct consequence of Dirichlet's test, we immediately get the alternating series test.

Theorem 6.5 (**Alternating series test**). *If $\{a_n\}$ is a sequence of bounded variation that converges to zero, then the sum $\sum(-1)^{n-1}a_n$ converges. In particular, if $\{a_n\}$ is a monotone sequence of real numbers approaching zero, then the sum $\sum(-1)^{n-1}a_n$ converges.*

Proof. Since the partial sums of $\sum(-1)^{n-1}$ are bounded and $\{a_n\}$ is of bounded variation and converges to zero, the sum $\sum(-1)^{n-1}a_n$ converges by Dirichlet's test. □

Example 6.5. The **alternating harmonic series**

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+-\cdots$$

converges. Of course, we already knew this and we also know that the value of the alternating harmonic series equals $\log 2$ (see Section 4.6).

We now come to a very useful theorem for approximation purposes.

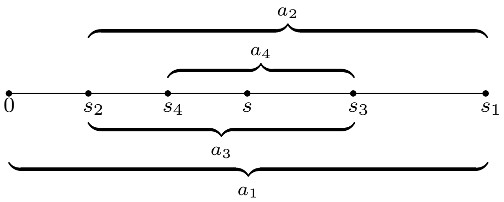


FIGURE 6.1. The partial sums $\{s_n\}$ jump forward and backward by the amounts given by the a_n's. This picture also shows that $|s - s_1| \le a_2$, $|s - s_2| \le a_3$, $|s - s_3| \le a_4$, ...

COROLLARY 6.6 (**Alternating series error estimate**). *If $\{a_n\}$ is a monotone sequence of real numbers approaching zero, and if s denotes the sum $\sum(-1)^{n-1}a_n$ and s_n denotes the n-th partial sum, then*

$$|s - s_n| \le |a_{n+1}|.$$

PROOF. To establish the error estimate, we assume that $a_n \ge 0$ for each n, in which case we have $a_1 \ge a_2 \ge a_3 \ge a_4 \ge \cdots \ge 0$. (The case when $a_n \le 0$ is similar or can be derived from the present case by multiplying by -1.) Let's consider how $s = \sum_{n=1}^{\infty}(-1)^{n-1}a_n$ is approximated by the s_n's. Observe that $s_1 = a_1$ increases from $s_0 = 0$ by the amount a_1; $s_2 = a_1 - a_2 = s_1 - a_2$ decreases from s_1 by the amount a_2; $s_3 = a_1 - a_2 + a_3 = s_2 + a_3$ increases from s_2 by the amount a_3, and so on; see Figure 6.1 for a picture of what's going on here. Studying this figure also shows why $|s - s_n| \le a_{n+1}$ holds. For this reason, we shall leave the exact proof details to the diligent and interested reader! □

Example 6.6. Suppose that we wanted to find $\log 2$ to two decimal places (in base 10); that is, we want to find b_0, b_1, b_2 in the decimal expansion $\log 2 = b_0.b_1b_2$ where by the usual convention, b_2 is "rounded up" if $b_3 \ge 5$. We can determine these decimals by finding n such that s_n, the n-th partial sum of $\log 2 = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}$, satisfies

$$|\log 2 - s_n| < 0.005;$$

that is,

$$\log 2 - 0.005 < s_n < \log 2 + 0.005.$$

Can you see why these inequalities guarantee that s_n has a decimal expansion starting with $b_0.b_1b_2$? Any case, according to the alternating series error estimate, we can make this this inequality hold by choosing n such that

$$|a_{n+1}| = \frac{1}{n+1} < 0.005 \quad\Longrightarrow\quad 500 < n+1 \quad\Longrightarrow\quad n = 500 \ \text{ works.}$$

With about five hours of pencil and paper work (and ten coffee breaks ☺) we find that $s_{500} = \sum_{n=1}^{500}\frac{(-1)^n}{n} = 0.69$ to two decimal places. Thus, $\log 2 = 0.69$ to two decimal places. A lot of work just to get two decimal places!

Example 6.7. (**Irrationality of e, Proof II**) Another nice application of the alternating series error estimate (or rather its proof) is a simple proof that e is irrational, cf. [**180**], [**7**]. Indeed, on the contrary, let us assume that $e = m/n$ where

$m, n \in \mathbb{N}$. Then we can write

$$\frac{n}{m} = e^{-1} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \quad \Longrightarrow \quad \frac{n}{m} - \sum_{k=0}^{m} \frac{(-1)^k}{k!} = \sum_{k=m+1}^{\infty} \frac{(-1)^k}{k!}.$$

Multiplying both sides by $m!/(-1)^{m+1} = \pm m!$, we obtain

$$\pm\left(n\,(m-1)! - \sum_{k=0}^{m} (-1)^k \frac{m!}{k!}\right) = \sum_{k=m+1}^{\infty} \frac{(-1)^{k-m-1} m!}{k!} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} m!}{(m+k)!}. \tag{6.3}$$

For $0 \leq k \leq m$, $m!/k!$ is an integer (this is because $m! = 1 \cdot 2 \cdots k \cdot (k+1) \cdots m$ contains a factor of $k!$), therefore the left-hand side of (6.3) is an integer, say $s \in \mathbb{Z}$, so that $s = \sum_{k=1}^{\infty} (-1)^{k-1} a_k$ where $a_k = \frac{m!}{(m+k)!}$. Thus, as seen in Figure 6.1, we have

$$0 < s < a_1 = \frac{1}{m+1}.$$

Now recall that $m \in \mathbb{N}$, so $1/(m+1) \leq 1/2$. Thus, s is an integer strictly between 0 and 1/2; an obvious contradiction!

6.1.5. Abel's test for series. Now let's modify the sum $\sum_{n=1}^{\infty} \frac{e^{inx}}{n}$, say to the slightly more complicated version

$$\sum_{n=1}^{\infty} \left(1 + \frac{1}{n}\right)^n \frac{e^{inx}}{n}.$$

If we try to determine the convergence of this series using Dirichlet's test, we'll have to do some work, but if we're feeling a little lazy, we can use the following theorem, whose proof uses an "$\varepsilon/3$-trick."

Theorem 6.7 (**Abel's test for series**). *Suppose that $\sum a_n$ converges. Then for any sequence $\{b_n\}$ of bounded variation, the series $\sum a_n b_n$ converges.*

Proof. We shall apply Abel's lemma to establish that the sequence of partial sums for $\sum a_n b_n$ forms a Cauchy sequence, which implies that $\sum a_n b_n$ converges. For $m < n$, by Abel's lemma, we have

$$\sum_{k=m+1}^{n} a_k b_k = s_n b_n - s_m b_m - \sum_{k=m}^{n-1} s_k (b_{k+1} - b_k), \tag{6.4}$$

where s_n is the n-th partial sum of the series $\sum a_n$. Adding and subtracting $s := \sum a_n$ to s_k on the far right of (6.4), we find that

$$\begin{aligned} \sum_{k=m}^{n-1} s_k (b_{k+1} - b_k) &= \sum_{k=m}^{n-1} (s_k - s)(b_{k+1} - b_k) + s \sum_{k=m}^{n-1} (b_{k+1} - b_k) \\ &= \sum_{k=m}^{n-1} (s_k - s)(b_{k+1} - b_k) + s b_n - s b_m, \end{aligned}$$

since the sum telescoped. Replacing this into (6.4), we obtain

$$\sum_{k=m+1}^{n} a_k b_k = (s_n - s) b_n - (s_m - s) b_m - \sum_{k=m}^{n-1} (s_k - s)(b_{k+1} - b_k).$$

Let $\varepsilon > 0$. Since $\{b_n\}$ is of bounded variation, this sequence converges by Proposition 6.3, so in particular is bounded and therefore, since $s_n \to s$, we have

$(s_n - s)b_n \to 0$ and $(s_m - s)b_m \to 0$. Thus, we can choose N such that for $n, m > N$, we have $|(s_n - s)b_n| < \varepsilon/3$, $|(s_m - s)b_m| < \varepsilon/3$, and $|s_n - s| < \varepsilon/3$. Thus, for $N < m < n$, we have

$$\begin{aligned}\left|\sum_{k=m+1}^{n} a_k b_k\right| &\le |(s_n - s)b_n| + |(s_m - s)b_m| + \sum_{k=m}^{n-1} |(s_k - s)(b_{k+1} - b_k)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \sum_{k=m}^{n-1} |b_{k+1} - b_k|.\end{aligned}$$

Finally, since $\sum |b_{k+1} - b_k|$ converges, by the Cauchy criterion for series, the sum $\sum_{k=m}^{n-1} |b_{k+1} - b_k|$ can be made less than 1 for N chosen larger if necessary. Thus, for $N < m < n$, we have $|\sum_{k=m+1}^{n} a_k b_k| < \varepsilon$. This completes our proof. □

Example 6.8. Back to our discussion above, we can write

$$\sum_{n=1}^{\infty} \Big(1 + \frac{1}{n}\Big)^n \frac{e^{inx}}{n} = \sum a_n\, b_n,$$

where $a_n = \frac{e^{inx}}{n}$ and $b_n = (1 + \frac{1}{n})^n$. Since we already know that $\sum_{n=1}^{\infty} a_n$ converges and that $\{b_n\}$ is nondecreasing and bounded above (by e — see Section 3.3) and therefore is of bounded variation, Abel's test shows that the series $\sum a_n b_n$ converges.

Exercises 6.1.

1. Following Fredricks and Nelsen [**77**], we use summation by parts to derive neat identities for the Fibonacci numbers. Recall that the Fibonacci sequence $\{F_n\}$ is defined as $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \ge 2$.
   (a) Let $a_n = F_{n+1}$ and $b_n = 1$ in the summation by parts formula (see Theorem 6.1) to derive the identity:
   $$F_1 + F_2 + F_3 + \cdots + F_n = F_{n+2} - 1.$$
   (b) Let $a_n = b_n = F_n$ in the summation by parts formula to get
   $$F_1^2 + F_2^2 + F_3^2 + \cdots + F_n^2 = F_n F_{n+1}.$$
   (c) What a_n's and b_n's would you choose to derive the formulas:
   $$F_1 + F_3 + F_5 + \cdots + F_{2n-1} = F_{2n} \quad , \quad 1 + F_2 + F_4 + F_6 + \cdots + F_{2n} = F_{2n+1}?$$
2. Following Fort [**76**], we relate limits of arithmetic means to summation by parts.
   (a) Let $\{a_n\}, \{b_n\}$ be sequences of complex numbers and assume that $b_n \to 0$ and $\frac{1}{n}\sum_{k=1}^{n} k\,|b_{k+1} - b_k| \to 0$ as $n \to \infty$, and that for some constant C, we have $|\frac{1}{n}\sum_{k=1}^{n} a_k| \le C$ for all n. Prove that
   $$\frac{1}{n}\sum_{k=1}^{n} a_k b_k \to 0 \quad \text{as } n \to \infty.$$
   (b) Apply this result to $a_n = (-1)^{n-1} n$ and $b_n = 1/\sqrt{n}$ to prove that
   $$\frac{1}{n}\Big(\sqrt{1} - \sqrt{2} + \sqrt{3} - \sqrt{4} + \cdots + (-1)^{n-1}\sqrt{n}\Big) = \frac{1}{n}\sum_{k=1}^{n} (-1)^k \sqrt{k} \to 0 \quad \text{as } n \to \infty.$$
3. Determine the convergence or divergence of the following series:
   $$(a)\ \frac{1}{1} + \frac{1}{2} + \frac{1}{3} - \frac{1}{4} - \frac{1}{5} + \frac{1}{6} + \frac{1}{7} - - + + \cdots \quad , \quad (b)\ \sum_{n=1}^{\infty} (-1)^n (\sqrt{n+1} - \sqrt{n}).$$

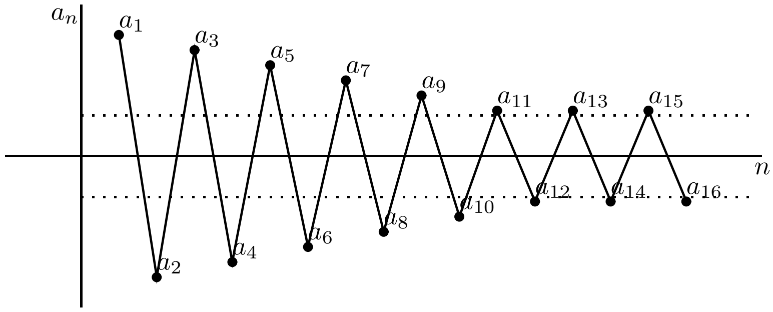


FIGURE 6.2. For the oscillating sequence $\{a_n\}$, the upper dashed line represents $\limsup a_n$ and the lower dashed line represents $\liminf a_n$.

$$(c)\ \sum_{n=2}^{\infty} \frac{\cos nx}{\log n} \quad , \quad (d)\ \frac{1}{2\cdot 1} - \frac{1}{2\cdot 2} + \frac{1}{3\cdot 3} - \frac{1}{3\cdot 4} + \frac{1}{4\cdot 5} - \frac{1}{4\cdot 6} + - \cdots$$

$$(e)\ \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \log \frac{2n+1}{n} \quad , \quad (f)\ \sum_{n=2}^{\infty} \cos nx \, \sin\Big(\frac{x}{n}\Big) \ (x \in \mathbb{R}) \quad , \quad (g)\ \sum_{n=2}^{\infty} (-1)^{n-1} \frac{\log n}{n}.$$

6.2. Liminfs/sups, ratio/roots, and power series

It is a fact of life that most sequences simply do not converge. In this section we introduce limit infimums and supremums, which *always* exist, either as real numbers or as $\pm\infty$. We also study their basic properties. We need these limits to study the ratio and root tests. You've probably seen these tests before in elementary calculus, but in this section we'll look at them in a slightly more sophisticated way.

6.2.1. Limit infimums and supremums. For an arbitrary sequence $\{a_n\}$ of real numbers we know that $\lim a_n$ may not exist; such as the sequence seen in Figure 6.2. However, being mathematicians we shouldn't let this stop up and in this subsection define "limits" for an arbitrary sequence. It turns out that there are two notions of "limit" that show up often, one is the limit supremum of $\{a_n\}$, which represents the "greatest" limiting value the a_n's could possibly have and the second is the limit infimum of $\{a_n\}$, which represents the "least" limiting value that the a_n's could possibly have. See Figure 6.2 for a picture of these ideas.

We now make "greatest" limiting value and "least" limiting value precise. Let $a_1, a_2, a_3, \ldots$ be any sequence of real numbers bounded from above. Let us put

$$s_n := \sup_{k \geq n} a_k = \sup\{a_n, a_{n+1}, a_{n+2}, a_{n+3}, \ldots\}.$$

Note that

$$s_{n+1} = \sup\{a_{n+1}, a_{n+2}, \ldots\} \leq \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} = s_n.$$

Indeed, s_n is an upper bound for $\{a_n, a_{n+1}, a_{n+2}, \ldots\}$ and hence an upper bound for $\{a_{n+1}, a_{n+2}, \ldots\}$, therefore s_{n+1}, being the least such upper bound, must satisfy $s_{n+1} \leq s_n$. Thus, $s_1 \geq s_2 \geq \cdots \geq s_n \geq s_{n+1} \geq \cdots$ is an nonincreasing sequence. In particular, being a monotone sequence, the limit $\lim s_n$ is defined either a real number or (properly divergent to) $-\infty$. We define

$$\boxed{\limsup a_n := \lim s_n = \lim_{n\to\infty}\Big(\sup\{a_n, a_{n+1}, a_{n+2}, \ldots\}\Big).}$$

This limit, which again is either a real number or $-\infty$, is called the **limit supremum** or **lim sup** of the sequence $\{a_n\}$. This name fits since $\limsup a_n$ is exactly that, a limit of supremums. If $\{a_n\}$ is not bounded from above, then we define

$$\boxed{\limsup a_n := \infty \quad \text{if } \{a_n\} \text{ is not bounded from above.}}$$

We define an **extended real number** as a real number or the symbols $\infty = +\infty$, $-\infty$. Then it is worth mentioning that lim sups *always* exist as an extended real number, unlike regular limits which may not exist. For the picture in Figure 6.2 notice that

$$\begin{aligned} s_1 &= \sup\{a_1, a_2, a_3, \ldots\} = a_1, \\ s_2 &= \sup\{a_2, a_3, a_4, \ldots\} = a_3, \\ s_3 &= \sup\{a_3, a_4, a_5, \ldots\} = a_3, \end{aligned}$$

and so on. Thus, the sequence $s_1, s_2, s_3, \ldots$ picks out the odd-indexed terms of the sequence $a_1, a_2, \ldots$. Therefore, $\limsup a_n = \lim s_n$ is the value given by the upper dashed line in Figure 6.2. (Of course, here we are assuming that the a_n's behave just as you think they should for for $n \geq 17$.) Here are some other examples.

Example 6.9. We shall compute $\limsup a_n$ where $a_n = \frac{1}{n}$. According to the definition of lim sup, we first have to find s_n:

$$s_n := \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} = \sup\left\{\frac{1}{n}, \frac{1}{n+1}, \frac{1}{n+2}, \frac{1}{n+3}, \ldots\right\} = \frac{1}{n}.$$

Second, we take the limit of the sequence $\{s_n\}$:

$$\limsup a_n := \lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{1}{n} = 0.$$

Notice that $\lim a_n$ also exists and $\lim a_n = 0$, the same as the $\limsup$. We'll come back to this observation in Example 6.11 below.

Example 6.10. Consider the sequence $\{(-1)^n\}$. In this case, we know that $\lim(-1)^n$ does not exist. To find $\limsup(-1)^n$, we first compute s_n:

$$s_n = \sup\{(-1)^n, (-1)^{n+1}, (-1)^{n+2}, \ldots\} = \sup\{+1, -1\} = 1,$$

where we used that the set $\{(-1)^n, (-1)^{n+1}, (-1)^{n+2}, \ldots\}$ is just a set consisting of the numbers $+1$ and -1. Hence,

$$\limsup(-1)^n := \lim s_n = \lim 1 = 1.$$

We can also define a corresponding $\liminf a_n$, which is a limit of infimums. To do so, assume for the moment that our generic sequence $\{a_n\}$ is bounded from below. Consider the sequence $\{\iota_n\}$ where

$$\iota_n := \inf_{k\geq n} a_k = \inf\{a_n, a_{n+1}, a_{n+2}, a_{n+3}, \ldots\}.$$

Note that

$$\iota_n = \inf\{a_n, a_{n+2}, \ldots\} \leq \inf\{a_{n+1}, a_{n+2}, \ldots\} = \iota_{n+1},$$

since the set $\{a_n, a_{n+2}, \ldots\}$ on the left of $\leq$ contains the set $\{a_{n+1}, a_{n+2}, \ldots\}$. Thus, $\iota_1 \leq \iota_2 \leq \cdots \leq \iota_n \leq \iota_{n+1} \leq \cdots$ is an nondecreasing sequence. In particular, being a monotone sequence, the limit $\lim \iota_n$ is defined either a real number or (properly divergent to) ∞. We define

$$\boxed{\liminf a_n := \lim \iota_n = \lim_{n\to\infty} \Big(\inf\{a_n, a_{n+1}, a_{n+2}, \ldots\} \Big),}$$

which exists either as a real number or $+\infty$, is called the **limit infimum** or **lim inf** of $\{a_n\}$. If $\{a_n\}$ is not bounded from below, then we define

$$\boxed{\liminf a_n := -\infty \quad \text{if } \{a_n\} \text{ is not bounded from below.}}$$

Again, as with lim sups, lim infs always exist as extended real numbers. For the picture in Figure 6.2 notice that

$$\begin{aligned}
\iota_1 &= \sup\{a_1, a_2, a_3, \ldots\} = a_2,\\
\iota_2 &= \sup\{a_2, a_3, a_4, \ldots\} = a_2,\\
\iota_3 &= \sup\{a_3, a_4, a_5, \ldots\} = a_4,
\end{aligned}$$

and so on. Thus, the sequence $\iota_1, \iota_2, \iota_3, \ldots$ picks out the even-indexed terms of the sequence $a_1, a_2, \ldots$. Therefore, $\limsup a_n = \lim \iota_n$ is the value given by the lower dashed line in Figure 6.2. Here are some more examples.

Example 6.11. We shall compute $\liminf a_n$ where $a_n = \frac{1}{n}$. According to the definition of $\liminf$, we first have to find ι_n:

$$\iota_n := \inf\{a_n, a_{n+1}, a_{n+2}, \ldots\} = \inf\left\{\frac{1}{n}, \frac{1}{n+1}, \frac{1}{n+2}, \frac{1}{n+3}, \ldots\right\} = 0.$$

Second, we take the limit of ι_n:

$$\liminf a_n := \lim_{n\to\infty} \iota_n = \lim_{n\to\infty} 0 = 0.$$

Notice that $\lim a_n$ also exists and $\lim a_n = 0$, the same as $\liminf a_n$, which is the same as $\limsup a_n$ as we saw in Example 6.9. We are thus lead to make the following conjecture: If $\lim a_n$ exists, then $\limsup a_n = \liminf a_n = \lim a_n$; this conjecture is indeed true as we'll see in Property *(2)* of Theorem 6.8.

Example 6.12. If $a_n = (-1)^n$, then

$$\inf\{a_n, a_{n+1}, a_{n+2}, \ldots\} = \sup\{(-1)^n, (-1)^{n+1}, (-1)^{n+2}, \ldots\} = \inf\{+1, -1\} = -1.$$

Hence,

$$\liminf (-1)^n := \lim -1 = -1.$$

The following theorem contains the main properties of limit infimums and supremums that we shall need in the sequel.

THEOREM 6.8 (**Properties of lim inf/sup**). *If $\{a_n\}$ and $\{b_n\}$ are sequences of real numbers, then*

(1) $\limsup a_n = -\liminf(-a_n)$ *and* $\liminf a_n = -\limsup(-a_n)$.

(2) $\lim a_n$ *is defined, as a real number or $\pm\infty$, if and only if* $\limsup a_n = \liminf a_n$*, in which case,*

$$\lim a_n = \limsup a_n = \liminf a_n.$$

(3) If $a_n \le b_n$ for all n sufficiently large, then

$$\liminf a_n \le \liminf b_n \quad \textit{and} \quad \limsup a_n \le \limsup b_n.$$

(4) The following inequality properties hold:

(a) $\limsup a_n < a \quad \Longrightarrow \quad$ *there is an N such that $n > N \Longrightarrow a_n < a$.*
(b) $\limsup a_n > a \quad \Longrightarrow \quad$ *there exist infinitely many n's such that $a_n > a$.*
(c) $\liminf a_n < a \quad \Longrightarrow \quad$ *there exist infinitely many n's such that $a_n < a$.*
(d) $\liminf a_n > a \quad \Longrightarrow \quad$ *there is an N such that $n > N \Longrightarrow a_n > a$.*

Proof. To prove *(1)* assume first that $\{a_n\}$ is not bounded from above; then $\{-a_n\}$ is not bounded from below. Hence, $\limsup a_n := \infty$ and $\liminf(-a_n) := -\infty$, which implies *(1)* in this case. Assume now that $\{a_n\}$ is bounded above. Recall from Lemma 2.29 that given any nonempty subset $A \subseteq \mathbb{R}$ bounded above, we have $\sup A = -\inf(-A)$. Hence,

$$\sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} = -\inf\{-a_n, -a_{n+1}, -a_{n+2}, -a_{n+3}, \ldots\}.$$

Taking $n \to \infty$ on both sides, we get $\limsup a_n = -\liminf(-a_n)$.

We now prove *(2)*. Suppose first that $\lim a_n$ converges to a real number L. Then given $\varepsilon > 0$, there exists an N such that

$$L - \varepsilon \le a_k \le L + \varepsilon, \quad \text{for all } k > N,$$

which implies that for any $n > N$,

$$L - \varepsilon \le \inf_{k \ge n} a_k \le \sup_{k \ge n} a_k \le L + \varepsilon.$$

Taking $n \to \infty$ implies that

$$L - \varepsilon \le \liminf a_n \le \limsup a_n \le L + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\limsup a_n = L = \liminf a_n$. Reversing these steps, we leave you to show that if $\limsup a_n = L = \liminf a_n$, then $\{a_n\}$ converges to L. We now consider *(2)* in the case that $\lim a_n = +\infty$; the case where the limit is $-\infty$ is proved similarly. Then given any real number $M > 0$, there exists an N such that

$$n > N \quad \Longrightarrow \quad M \le a_n.$$

This implies that

$$M \le \inf_{k \ge n} a_k \le \sup_{k \ge n} a_k.$$

Taking $n \to \infty$ we obtain

$$M \le \liminf a_n \le \limsup a_n.$$

Since $M > 0$ was arbitrary, it follows that $\limsup a_n = +\infty = \liminf a_n$. Reversing these steps, we leave you to show that if $\limsup a_n = +\infty = \liminf a_n$, then $a_n \to +\infty$.

To prove *(3)* note that if $\{a_n\}$ is not bounded from below, then $\liminf a_n := -\infty$ so $\liminf a_n \le \liminf b_n$ automatically; thus, we may assume that $\{a_n\}$ is bounded from below. In this case, observe that $a_n \le b_n$ for all n sufficiently large implies that, for n sufficiently large,

$$\inf\{a_n, a_{n+1}, a_{n+2}, \ldots\} \le \inf\{b_n, b_{n+1}, b_{n+2}, b_{n+3}, \ldots\}$$

Taking $n \to \infty$, and using that limits preserve inequalities, now proves *(3)*. The proof that $\limsup a_n \le \limsup b_n$ is similar.

Because this proof is becoming unbearably unbearable ☺ we'll only prove *(a)*, *(b)* of *(4)* leaving *(c)*, *(d)* to the reader. Assume that $\limsup a_n < a$, that is,

$$\lim_{n\to\infty} \Big(\sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} \Big) < a.$$

It follows that for some N, we have

$$n > N \quad \Longrightarrow \quad \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} < a,$$

that is, the least upper bound of $\{a_n, a_{n+1}, a_{n+2}, \ldots\}$ is strictly less than a, so we must have we have $a_n < a$ for all $n > N$. Assume now that $\limsup a_n > a$. If $\{a_n\}$ is not bounded from above then there must exist infinitely many n's such that $a_n > a$, for otherwise if there were only finitely many n's such that $a_n > a$, then $\{a_n\}$ would be bounded from above. Assume now that $\{a_n\}$ is bounded from above. Then,

$$\lim_{n\to\infty} \Big(\sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} \Big) > a$$

implies that for some N, we have

$$n > N \quad \Longrightarrow \quad \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} > a.$$

Now if there were only finitely many n's such that $a_n > a$, then we can choose $n > N$ large enough such that $a_k \leq a$ for all $k \geq n$. However, this would imply that for such n, $\sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} \leq a$, a contradiction. Hence, there are infinitely many n's such that $a_n > a$. □

6.2.2. Ratio/root tests, and the exponential and ζ-functions, again. In elementary calculus you should have studied the ratio test: If the limit $L_1 := \lim \left|\frac{a_{n+1}}{a_n}\right|$ *exists*, then the series $\sum a_n$ converges if $L_1 < 1$ and diverges if $L_1 > 1$ (if $L_1 = 1$, then the test is inconclusive). You also studied the root test: If the limit $L_2 := \lim |a_n|^{1/n}$ *exists*, then the series $\sum a_n$ converges if $L_2 < 1$ and diverges if $L_2 > 1$ (if $L_2 = 1$, then the test is inconclusive). Now what if the limits $\lim \left|\frac{a_{n+1}}{a_n}\right|$ or $\lim |a_n|^{1/n}$ don't exist, are there still ratio and root tests? The answer is "yes," but we have to replace lim with lim inf's and lim sup's. Before stating these new ratio/root tests, we first consider the following important lemma.

Lemma 6.9. *If $\{a_n\}$ is a sequence of nonzero complex numbers, then*

$$\liminf \left|\frac{a_{n+1}}{a_n}\right| \leq \liminf |a_n|^{1/n} \leq \limsup |a_n|^{1/n} \leq \limsup \left|\frac{a_{n+1}}{a_n}\right|.$$

Proof. The middle inequality is automatic (because inf's are $\leq$ sup's), so we just need to prove the left and right inequalities. Consider the left one; the right one is analogous and is left to the reader. If $\liminf |a_{n+1}/a_n| = -\infty$, then there is nothing to prove, so we may assume that $\liminf |a_{n+1}/a_n| \neq -\infty$. Given any $b < \liminf |a_{n+1}/a_n|$, we shall prove that $b < \liminf |a_n|^{1/n}$. This proves the left side in our desired inequalities, for, if on the contrary we have $\liminf |a_n|^{1/n} < \liminf |a_{n+1}/a_n|$, then choosing $b = \liminf |a_n|^{1/n}$, we would have

$$\liminf |a_n|^{1/n} < \liminf |a_n|^{1/n},$$

an obvious contradiction. So, let $b < \liminf |a_{n+1}/a_n|$. Choose a such that $b < a < \liminf |a_{n+1}/a_n|$. Then by Property *4 (d)* in Theorem 6.8, for some N, we have

$$n > N \quad \Longrightarrow \quad \left|\frac{a_{n+1}}{a_n}\right| > a.$$

Fix $m > N$ and let $n > m > N$. Then we can write

$$|a_n| = \left|\frac{a_n}{a_{n-1}}\right| \cdot \left|\frac{a_{n-1}}{a_{n-2}}\right| \cdots \left|\frac{a_{m+1}}{a_m}\right| \cdot |a_m|.$$

There are $n - m$ quotients in this equality, each of which is greater than a, so

$$|a_n| > a \cdot a \cdots a \cdot |a_m| = a^{n-m} \cdot |a_m|,$$

which implies that

$$|a_n|^{1/n} > a^{1-m/n} \cdot |a_m|^{1/n}. \tag{6.5}$$

Since

$$\lim_{n\to\infty} a^{1-m/n} \cdot |a_m|^{1/n} = a,$$

and limit infimums preserve inequalities, we have

$$\liminf |a_n|^{1/n} \geq \liminf a^{1-m/n} \cdot |a_m|^{1/n} = \lim_{n\to\infty} a^{1-m/n} \cdot |a_m|^{1/n} = a,$$

where we used Property *(2)* of Theorem 6.8. Since $a > b$, we have $b < \liminf |a_n|^{1/n}$ and our proof is complete. □

Here's Cauchy's root test, a far-reaching generalization of the root test you learned in elementary calculus.

THEOREM 6.10 (**Cauchy's root test**). *A series $\sum a_n$ converges absolutely or diverges according as*

$$\limsup \left|a_n\right|^{1/n} < 1 \qquad \textit{or} \qquad \limsup \left|a_n\right|^{1/n} > 1.$$

PROOF. Suppose first that $\limsup \left|a_n\right|^{1/n} < 1$. Then we can choose $0 < a < 1$ such that $\limsup \left|a_n\right|^{1/n} < a$, which, by Property *4 (a)* of Theorem 6.8, implies that for some N,

$$n > N \quad \Longrightarrow \quad \left|a_n\right|^{1/n} < a,$$

that is,

$$n > N \quad \Longrightarrow \quad \left|a_n\right| < a^n.$$

Since $a < 1$, we know that the infinite series $\sum a^n$ converges; thus by the comparison test, the sum $\sum |a_n|$ also converges, and hence $\sum a_n$ converges as well.

Assume now that $\limsup \left|a_n\right|^{1/n} > 1$. Then by Property *4 (b)* of Theorem 6.8, there are infinitely many n's such that $\left|a_n\right|^{1/n} > 1$. Thus, there are infinitely many n's such that $|a_n| > 1$. Hence by the n-th term test, the series $\sum a_n$ cannot converge. □

It is important to remark that in the other case, that is, $\limsup \left|a_n\right|^{1/n} = 1$, this test does not give information as to convergence.

Example 6.13. Consider the series $\sum 1/n$, which diverges, and observe that $\limsup |1/n|^{1/n} = \lim 1/n^{1/n} = 1$ (see Section 3.1 for the proof that $\lim n^{1/n} = 1$). However, $\sum 1/n^2$ converges, and $\limsup |1/n^2|^{1/n} = \lim(1/n^{1/n})^2 = 1$ as well, so when $\limsup |a_n|^{1/n} = 1$ it's not possible to tell whether or not the series converges.

As with the root test, in elementary calculus you learned the ratio test most likely without proof, and, accepting by faith this test as correct you probably used it to determine the convergence/divergence of many types of series. Here's d'Alembert's ratio test, a far-reaching generalization of the ratio test[2].

[2] *Allez en avant, et la foi vous viendra [push on and faith will catch up with you]. Advice to those who questioned the calculus by Jean Le Rond d'Alembert (1717–1783)* [**141**]

THEOREM 6.11 (**d'Alembert's ratio test**). *A series $\sum a_n$, with a_n nonzero for n sufficiently large, converges absolutely or diverges according as*

$$\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1 \qquad \textit{or} \qquad \liminf \left|\frac{a_{n+1}}{a_n}\right| > 1.$$

PROOF. If we set $L := \limsup |a_n|^{1/n}$, then by Lemma 6.9, we have

$$\liminf \left|\frac{a_{n+1}}{a_n}\right| \leq L \leq \limsup \left|\frac{a_{n+1}}{a_n}\right|. \tag{6.6}$$

Therefore, if $\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$, then $L < 1$ too, so $\sum a_n$ converges absolutely by the root test. On the other hand, if $\liminf \left|\frac{a_{n+1}}{a_n}\right| > 1$, then $L > 1$ too, so $\sum a_n$ diverges by the root test. □

We remark that in the other case, that is, $\liminf \left|\frac{a_{n+1}}{a_n}\right| \leq 1 \leq \limsup \left|\frac{a_{n+1}}{a_n}\right|$, this test does not give information as to convergence. Indeed, the same divergent and convergent examples used for the root test, $\sum 1/n$ and $\sum 1/n^2$, have the property that $\liminf \left|\frac{a_{n+1}}{a_n}\right| = 1 = \limsup \left|\frac{a_{n+1}}{a_n}\right|$.

Note that if $\limsup |a_n|^{1/n} = 1$, that is, the root test fails (to give a decisive answer), then setting $L = 1$ in (6.6), we see that the ratio test also fails. Thus,

$$\text{root test fails} \implies \text{ratio test fails.} \tag{6.7}$$

Therefore, if the root test fails one cannot hope to appeal to the ratio test.

Let's now consider some examples.

Example 6.14. First, our old friend:

$$\exp(z) := \sum_{n=1}^{\infty} \frac{z^n}{n!},$$

which we already knows converges, but for the fun of it, let's apply the ratio test. Observe that

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{\frac{z^{n+1}}{(n+1)!}}{\frac{z^n}{n!}}\right| = |z| \cdot \frac{n!}{(n+1)!} = \frac{|z|}{n+1}.$$

Hence,

$$\lim \left|\frac{a_{n+1}}{a_n}\right| = 0 < 1.$$

Thus, the exponential function $\exp(z)$ converges absolutely for all $z \in \mathbb{C}$. This proof was a little easier than the one in Section 3.7, but then again, back then we didn't have the up-to-day technology of the ratio test that we have now. Here's an example that fails.

Example 6.15. Consider the Riemann zeta function

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}, \qquad \operatorname{Re} z > 1.$$

If $z = x + iy$ is separated into its real and imaginary parts, then

$$|a_n|^{1/n} = \left|\frac{1}{n^z}\right|^{1/n} = \Big(\frac{1}{n^x}\Big)^{1/n} = \Big(\frac{1}{n^{1/n}}\Big)^x.$$

Since $\lim n^{1/n} = 1$, it follows that

$$\lim \bigl|a_n\bigr|^{1/n} = 1$$

so the root test fails to give information, which also implies that the ratio test fails as well. Of course, using the comparison test as we did in the proof of Theorem 4.33 we already know that $\zeta(z)$ converges for all $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$.

It's easy to find examples of series for which the ratio test fails but the root test succeeds.

Example 6.16. A general class of examples that foil the ratio test are (see Problem 4)

$$a + b + a^2 + b^2 + a^3 + b^3 + a^4 + b^4 + \cdots \quad , \quad 0 < b < a < 1; \tag{6.8}$$

here, the odd terms are given by $a_{2n-1} = a^n$ and the even terms are given by $a_{2n} = b^n$. For concreteness, let us consider the series

$$\frac{1}{2} + \frac{1}{3} + \Bigl(\frac{1}{2}\Bigr)^2 + \Bigl(\frac{1}{3}\Bigr)^2 + \Bigl(\frac{1}{2}\Bigr)^3 + \Bigl(\frac{1}{3}\Bigr)^3 + \Bigl(\frac{1}{2}\Bigr)^4 + \Bigl(\frac{1}{3}\Bigr)^4 + \cdots .$$

Since

$$\Bigl|\frac{a_{2n}}{a_{2n-1}}\Bigr| = \Bigl|\frac{(1/3)^n}{(1/2)^n}\Bigr| = \Bigl(\frac{2}{3}\Bigr)^n$$

and

$$\Bigl|\frac{a_{2n+1}}{a_{2n}}\Bigr| = \Bigl|\frac{(1/2)^{n+1}}{(1/3)^n}\Bigr| = \Bigl(\frac{3}{2}\Bigr)^n \cdot \frac{1}{2},$$

It follows that $\liminf |a_{n+1}/a_n| = 0 < 1 < \infty = \limsup |a_{n+1}/a_n|$, so the ratio test does not give information. On the other hand, since

$$|a_{2n-1}|^{1/(2n-1)} = \bigl((1/2)^n\bigr)^{1/(2n-1)} = \Bigl(\frac{1}{2}\Bigr)^{\frac{n}{2n-1}}$$

and

$$|a_{2n}|^{1/(2n)} = \bigl((1/3)^{n-1}\bigr)^{1/(2n)} = \Bigl(\frac{1}{3}\Bigr)^{\frac{n-1}{2n}}$$

we leave it as an exercise for you to show that $\limsup |a_n|^{1/n} = (1/2)^{1/2}$. Since $(1/2)^{1/2} < 1$, the series converges by the root test.

Thus, in contrast to (6.7),

$$\text{ratio test fails} \quad \not\Longrightarrow \quad \text{root test fails.}$$

However, in the following lemma we show that if the ratio test fails such that the *true* limit $\lim |\frac{a_{n+1}}{a_n}| = 1$, then the root test fails as well.

Lemma 6.12. *If* $|\frac{a_{n+1}}{a_n}| \to L$ *with* L *an extended real number, then* $|a_n|^{1/n} \to L$.

Proof. By Lemma 6.9, we know that

$$\liminf \Bigl|\frac{a_{n+1}}{a_n}\Bigr| \le \liminf \bigl|a_n\bigr|^{1/n} \le \limsup \bigl|a_n\bigr|^{1/n} \le \limsup \Bigl|\frac{a_{n+1}}{a_n}\Bigr|.$$

By Theorem 6.8, a limit exists if and only if the lim inf and the lim sup have the same limit, so the outside quantities in these inequalities equal L. It follows that $\liminf |a_n|^{1/n} = \limsup |a_n|^{1/n} = L$ as well, and hence $\lim |a_n|^{1/n} = L$. □

Let's do one last (important) example:

Example 6.17. Consider the series

$$1+\sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots(2n)\,(2n+1)}. \tag{6.9}$$

Applying the ratio test, we have

$$\frac{a_{n+1}}{a_n}=\frac{(2n+1)(2n+1)}{(2n+2)(2n+3)}=\frac{4n^2+8n+1}{4n^2+10n+6}=\frac{1+\dfrac{2}{n}+\dfrac{1}{4n^2}}{1+\dfrac{5}{2n}+\dfrac{3}{2n^2}}. \tag{6.10}$$

Therefore, $\lim|\frac{a_{n+1}}{a_n}|=1$, so the ratio and root test give no information! What can we do? We'll see that Raabe's test in Section 6.3 will show that (6.9) converges.

6.2.3. Power series.

Our old friend

$$\exp(z):=\sum_{n=0}^{\infty}\frac{z^n}{n!}$$

is an example of a **power series**, by which we mean a series of the form

$$\sum_{n=0}^{\infty}a_n\,z^n,\quad \text{where}\quad z\in\mathbb{C},\qquad\text{or}\qquad \sum_{n=0}^{\infty}a_n\,x^n,\quad\text{where}\quad x\in\mathbb{R},$$

where $a_n\in\mathbb{C}$ for all n (in particular, the a_n's may be real). However, we shall focus on power series of the complex variable z *although essentially everything we mention works for real variables* x.

Example 6.18. Besides the exponential function, other familiar examples of power series include the trigonometric series, $\sin z=\sum_{n=0}^{\infty}(-1)^n z^{2n+1}/(2n+1)!$, $\cos z=\sum_{n=0}^{\infty}(-1)^n z^{2n}/(2n)!$.

The convergence of power series is quite easy to analyze. First, $\sum_{n=0}^{\infty}a_n z^n = a_0+a_1z+a_2z^2+\cdots$ certainly converges if $z=0$. For $|z|>0$ we can use the root test: Observe that (see Problem 8 for the proof that we can take out $|z|$)

$$\limsup\left|a_nz^n\right|^{1/n}=\limsup\left(|z|\,|a_n|^{1/n}\right)=|z|\,\limsup|a_n|^{1/n}.$$

Therefore, $\sum a_nz^n$ converges (absolutely) or diverges according as

$$|z|\cdot\limsup|a_n|^{1/n}<1\quad\text{or}\quad|z|\cdot\limsup|a_n|^{1/n}>1.$$

Therefore, if we define $0\le R\le\infty$ by

$$\boxed{R:=\frac{1}{\limsup|a_n|^{1/n}}} \tag{6.11}$$

where by convention, we put $R:=+\infty$ when $\limsup|a_n|^{1/n}=0$ and $R:=0$ when $\limsup|a_n|^{1/n}=+\infty$, then it follows that $\sum a_nz^n$ converges (absolutely) or diverges according to $|z|<R$ or $|z|>R$; when $|z|=R$, anything can happen. According to Figure 6.3, it is quite fitting to call R the **radius of convergence**. Let us summarize our findings in the following theorem named after Cauchy (whom we've already met many times) and Jacques Hadamard (1865–1963).[3]

[3] *The shortest path between two truths in the real domain passes through the complex domain. Jacques Hadamard (1865–1963). Quoted in The Mathematical Intelligencer 13 (1991).*

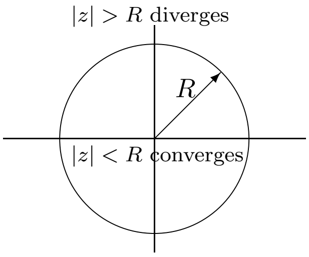


FIGURE 6.3. $\sum a_n z^n$ converges (absolutely) or diverges according as $|z| < R$ or $|z| > R$.

THEOREM 6.13 (**Cauchy-Hadamard theorem**). *If R is the radius of convergence of the power series $\sum a_n z^n$, then the series is absolutely convergent for $|z| < R$ and is divergent for $|z| > R$.*

One final remark. Suppose that the a_n's are nonzero for n sufficiently large and $\lim |\frac{a_n}{a_{n+1}}|$ exists. Then by Lemma 6.12, we have

$$\boxed{R = \lim \left| \frac{a_n}{a_{n+1}} \right|}. \tag{6.12}$$

This formula for the radius of convergence might, in some cases, be easier to work with than the formula involving $|a_n|^{1/n}$.

EXERCISES 6.2.

1. Find the lim inf/sups of the sequence $\{a_n\}$, where a_n is given by

$$(a)\ \frac{2+(-1)^n}{4}\ ,\ \ (b)\ (-1)^n\Big(1-\frac{1}{n}\Big)\ ,\ \ (c)\ 2^{(-1)^n}\ ,\ \ (d)\ 2^{n(-1)^n}\ ,\ \ (e)\ \Big(1+\frac{(-1)^n}{2}\Big)^n.$$

(f) If $\{r_n\}$ is a list of all rationals in $(0,1)$, prove $\liminf r_n = 0$ and $\limsup r_n = 1$.

2. Investigate the following series for convergence (in (c), $z \in \mathbb{C}$):

$$(a)\ \sum_{n=1}^{\infty} \frac{(n+1)(n+2)\cdots(n+n)}{n^n}\ ,\ (b)\ \sum_{n=1}^{\infty} \frac{(n+1)^n}{n!}\ ,\ (c)\ \sum_{n=1}^{\infty} \frac{n^z}{n!}\ ,\ (d)\ \sum_{n=1}^{\infty} \frac{1}{2^{n+(-1)^n}}.$$

3. Determine the radius of convergence for the following series:

$$(a)\ \sum_{n=1}^{\infty} \frac{(n+1)^n}{n^{n+1}} z^n\ \ ,\ \ (b)\ \sum_{n=1}^{\infty} \Big(\frac{n}{n+1}\Big)^n z^n\ \ ,\ \ (c)\ \sum_{n=1}^{\infty} \frac{(2n)!}{(n!)^2} z^n\ \ ,\ \ (d)\ \sum_{n=1}^{\infty} \frac{z^n}{n^p},$$

where in the last sum, $p \in \mathbb{R}$. If $z = x \in \mathbb{R}$, state all $x \in \mathbb{R}$ such that the series converge. For (c), your answer should depend on p.

4. (a) Investigate the series (6.8) for convergence using both the ratio and the root tests.
(b) Here is another class of examples:

$$1 + a + b^2 + a^3 + b^4 + a^5 + b^6 + \cdots \quad , \quad 0 < a < b < 1.$$

Show that the ratio test fails but the root test works.

5. Lemma 6.12 is very useful to determine certain limits which aren't obvious at first glance. Using this lemma, derive the following limits:

$$(a)\ \lim \frac{n}{(n!)^{1/n}} = e\ \ ,\ \ (b)\ \lim \frac{n+1}{(n!)^{1/n}} = e\ \ ,\ \ (c)\ \lim \frac{n}{[(n+1)(n+2)\cdots(n+n)]^{1/n}} = \frac{e}{4},$$

and for $a, b \in \mathbb{R}$ with $a > 0$ and $a + b > 0$,

$$(d)\ \lim \frac{n}{[(a+b)(2a+b)\cdots(na+b)]^{1/n}} = \frac{e}{a}.$$

Suggestion: For (a), let $a_n = n^n/n!$. Prove that $\lim \frac{a_{n+1}}{a_n} = e$ and hence $\lim a_n^{1/n} = e$ as well. As a side remark, recall that (a) is called (the "weak") Stirling's formula, which we introduced in (3.29) and proved in Problem 5 of Exercises 3.3.

6. In this problem we investigate the interesting power series $\sum_{n=1}^\infty \frac{n!}{n^n} z^n$, where $z \in \mathbb{C}$.
   (a) Prove that this series has radius of convergence $R = e$.
   (b) If $|z| = e$, then the ratio and root test both fail. However, if $|z| = e$, then prove that the infinite series diverges.
   (c) Investigate the convergence/divergence of $\sum_{n=1}^\infty \frac{n^n}{n!} z^n$, where $z \in \mathbb{C}$.
7. In this problem we investigate the interesting power series

$$F(z) := \sum_{n=0}^\infty F_{n+1}\, z^n = F_1 + F_2 z + F_3 z^2 + \cdots ,$$

where $\{F_n\}$ is the **Fibonacci sequence** defined in Problem 9 of Exercises 2.2: $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. In that problem you proved that $F_n = \frac{1}{\sqrt{5}}[\Phi^n - (-\Phi)^{-n}]$ where $\Phi = \frac{1+\sqrt{5}}{2}$, the **Golden ratio**.
   (i) Prove that $F(z)$ has radius of convergence equal to Φ^{-1}.
   (ii) Prove that for all z with $|z| < \Phi^{-1}$, we have $F(z) = \frac{1}{1-z-z^2}$. Suggestion: Show that $(1 - z - z^2)F(z) = 1$. By the way, given any sequence $\{a_n\}_{n=0}^\infty$, the power series $\sum_{n=0}^\infty a_n z^n$ is called the **generating function** of the sequence $\{a_n\}$. Thus, the generating function for $\{F_{n+1}\}$ has the closed form $1/(1 - z - z^2)$. For more on generating functions, see the free book [**246**]. Also, if you're interested in a magic trick you can do with the formula $F(z) = 1/(1 - z - z^2)$, see [**176**].
8. Here are some lim inf/sup problems. Let $\{a_n\}, \{b_n\}$ be sequences of real numbers.
   (a) Prove that if $c > 0$, then $\liminf(ca_n) = c \liminf a_n$ and $\limsup(ca_n) = c \limsup a_n$. Here, we take the "obvious" conventions: $c \cdot \pm\infty = \pm\infty$.
   (b) Prove that if $c < 0$, then $\liminf(ca_n) = c \limsup a_n$ and $\limsup(ca_n) = c \liminf a_n$.
   (c) If $\{a_n\}, \{b_n\}$ are bounded, prove that $\liminf a_n + \liminf b_n \leq \liminf(a_n + b_n)$.
   (d) If $\{a_n\}, \{b_n\}$ are bounded, prove that $\limsup(a_n + b_n) \leq \limsup a_n + \limsup b_n$.
9. If $a_n \to L$ where L is a positive real number, prove that $\limsup(a_n \cdot b_n) = L \limsup b_n$ and $\liminf(a_n \cdot b_n) = L \liminf b_n$. Here are some steps if you want them:
   (i) Show that you can get the lim inf statement from the lim sup statement, hence we can focus on the lim sup statement. We shall prove that $\limsup(a_n b_n) \leq L \limsup b_n$ and $L \limsup b_n \leq \limsup(a_n b_n)$.
   (ii) Show that the inequality $\limsup(a_n b_n) \leq L \limsup b_n$ follows if the following statement holds: If $\limsup b_n < b$, then $\limsup(a_n b_n) < L\, b$.
   (iii) Now prove that if $\limsup b_n < b$, then $\limsup(a_n b_n) < L\, b$. Suggestion: If $\limsup b_n < b$, then choose a such that $\limsup b_n < a < b$. Using Property *4 (a)* of Theorem 6.8 and the definition of $L = \lim a_n > 0$, prove that there is an N such that $n > N$ implies $b_n < a$ and $a_n > 0$. Conclude that for $n > N$, $a_n b_n < a a_n$. Finally, take lim sups of both sides of $a_n b_n < a a_n$.
   (iv) Show that the inequality $L \limsup b_n \leq \limsup(a_n b_n)$ follows if the following statement holds: If $\limsup(a_n b_n) < L\, b$, then $\limsup b_n < b$; then prove this statement.
10. Let $\{a_n\}$ be a sequence of real numbers. We prove that there are monotone subsequences of $\{a_n\}$ that converge to $\liminf a_n$ and $\limsup a_n$. Proceed as follows:
   (i) Using Theorem 3.13, show that it suffices to prove that there are subsequences converging to $\liminf a_n$ and $\limsup a_n$
   (ii) Show that it suffices to that there is a subsequence converging to $\liminf a_n$.
   (iii) If $\liminf a_n = \pm\infty$, prove there is a subsequence converging to $\liminf a_n$.
   (iv) Now assume that $\liminf a_n = \lim_{n\to\infty} \big(\inf\{a_n, a_{n+1}, \ldots\}\big) \in \mathbb{R}$. By definition of limit, show that there is an n so that $a - 1 < \inf\{a_n, a_{n+1}, \ldots\} < a + 1$. Show

that we can choose an n_1 so that $a - 1 < a_{n_1} < a + 1$. Then show there an $n_2 > n_1$ so that $a - \frac{1}{2} < a_{n_2} < a + \frac{1}{2}$. Continue this process.

6.3. A potpourri of ratio-type tests and "big $\mathcal{O}$" notation

In the previous section, we left it in the air whether or not the series

$$1 + \sum_{n=1}^{\infty} \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)\,(2n+1)}$$

converges (both the ratio and root tests failed). In this section we'll develop some new technologies that are able to detect the convergence of this series and other series for which the ratio and root tests *fail* to give information.

6.3.1. Kummer's test. The fundamental enhanced version of the ratio test is named after Ernst Kummer (1810–1893), from which we'll derive a potpourri of other ratio-type tests.

Theorem 6.14 (**Kummer's test**). *Let $\{a_n\}$ and $\{b_n\}$ be sequences of positive numbers where the sum $\sum b_n$ diverges, and define*

$$\kappa_n = \frac{1}{b_n}\frac{a_n}{a_{n+1}} - \frac{1}{b_{n+1}}.$$

Then $\sum a_n$ converges or diverges according as $\liminf \kappa_n > 0$ or $\limsup \kappa_n < 0$. In particular, if κ_n tends to some definite limit, κ, then $\sum a_n$ converges to diverges according as $\kappa > 0$ or $\kappa < 0$.

Proof. If $\liminf \kappa_n > 0$, then by Property *4 (d)* of Theorem 6.8, given any positive number a less than this limit infimum, there is an N such that

$$n > N \quad \Longrightarrow \quad \frac{1}{b_n}\frac{a_n}{a_{n+1}} - \frac{1}{b_{n+1}} > a.$$

Thus,

$$n > N \quad \Longrightarrow \quad \frac{1}{b_n}a_n - \frac{1}{b_{n+1}}a_{n+1} > a\,a_{n+1}. \tag{6.13}$$

Let $m > N$ and let $n > m > N$. Then (6.13) implies that

$$\sum_{k=m}^{n} a\,a_{k+1} < \sum_{k=m}^{n}\left(\frac{1}{b_k}a_k - \frac{1}{b_{k+1}}a_{k+1}\right) = \frac{1}{b_m}a_m - \frac{1}{b_{n+1}}a_{n+1},$$

since the sum telescoped. Therefore, as $\frac{1}{b_{n+1}}a_{n+1} > 0$, we have $\sum_{k=m}^{n} a\,a_{k+1} < \frac{1}{b_m}a_m$, or more succinctly,

$$\sum_{k=m}^{n} a_{k+1} < C$$

where $C = \frac{1}{a}\frac{1}{b_m}a_m$ is a constant independent of n. Since $n > m$ is completely arbitrary it follows that the partial sums of $\sum a_n$ always remain bounded by a fixed constant, so the sum must converge.

Assume now that $\limsup \kappa_n < 0$. Then by property *4 (a)* of Theorem 6.8, there is an N such that for all $n > N$, $\kappa_n < 0$, that is,

$$n > N \quad \Longrightarrow \quad \frac{1}{b_n}\frac{a_n}{a_{n+1}} - \frac{1}{b_{n+1}} < 0, \quad \text{that is,} \quad \frac{a_n}{b_n} < \frac{a_{n+1}}{b_{n+1}}.$$

Thus, for $n > N$, $\frac{a_n}{b_n}$ is increasing with n. In particular, fixing $m > N$, for all $n > m$, we have $C < a_n/b_n$, where $C = a_m/b_m$ is a constant independent of n. Thus, for all $n > m$, we have $Cb_n < a_n$ and since the sum $\sum b_n$ diverges, the comparison test implies that $\sum a_n$ diverges too. □

Note that d'Alembert's ratio test is just Kummer's test with $b_n = 1$ for each n.

6.3.2. Raabe's test and "big $\mathcal{O}$" notation. The following test, attributed to Joseph Ludwig Raabe (1801–1859), is just Kummer's test with the b_n's making up the harmonic series: $b_n = 1/n$.

THEOREM 6.15 (**Raabe's test**). *A series $\sum a_n$ of positive terms converges or diverges according as*

$$\liminf \ n\left(\frac{a_n}{a_{n+1}} - 1\right) > 0 \qquad \textit{or} \qquad \limsup \ n\left(\frac{a_n}{a_{n+1}} - 1\right) < 0.$$

In order to effectively apply Raabe's test, it is useful to first introduce some very handy notation. For a nonnegative function g, when we write $f = \mathcal{O}(g)$ ("**big O**" of g), we simply mean that $|f| \leq Cg$ for some constant C. In words, the big $\mathcal{O}$ notation just represents "a function that is in absolute value less than or equal to a constant times". This big $\mathcal{O}$ notation was introduced by Paul Bachmann (1837–1920) but became well-known through Edmund Landau (1877–1938) [**239**].

Example 6.19. For $x \geq 0$, we have

$$\frac{x^2}{1+x} = \mathcal{O}(x^2)$$

because $x^2/(1+x) \leq x^2$ for $x \geq 0$. Thus, for $x \geq 0$,

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x} \quad \Longrightarrow \quad \frac{1}{1+x} = 1 - x + \mathcal{O}(x^2). \tag{6.14}$$

In this section, we are mostly interested in using the big $\mathcal{O}$ notation when dealing with natural numbers.

Example 6.20. For $n \in \mathbb{N}$,

$$\frac{2}{n} + \frac{1}{4n^2} = \mathcal{O}\left(\frac{1}{n}\right), \tag{6.15}$$

because $\frac{2}{n} + \frac{1}{4n^2} \leq \frac{2}{n} + \frac{1}{4n} = \frac{C}{n}$ where $C = 2 + 1/4 = 9/4$.

Three important properties of the big $\mathcal{O}$ notation are (1) if $f = \mathcal{O}(ag)$ with $a \geq 0$, then $f = \mathcal{O}(g)$, and if $f_1 = \mathcal{O}(g_1)$ and $f_2 = \mathcal{O}(g_2)$, then (2) $f_1 f_2 = \mathcal{O}(g_1\, g_2)$ and (3) $f_1 + f_2 = \mathcal{O}(g_1 + g_2)$. To prove these properties, observe that if $|f| \leq C(ag)$, then $|f| \leq C'g$, where $C' = aC$, and that $|f_1| \leq C_1 g_1$ and $|f_2| \leq C_2 g_2$ imply

$$|f_1 f_2| \leq \big(C_1 C_2\big)\, g_1 g_2 \quad \text{and} \quad |f_1 + f_2| \leq (C_1 + C_2)\,(g_1 + g_2);$$

hence, our three properties.

Example 6.21. Thus, in view of (6.15), we have $\mathcal{O}\big(\frac{2}{n}+\frac{1}{4n^2}\big)^2 = \mathcal{O}\left(\frac{1}{n}\cdot\frac{1}{n}\right) = \mathcal{O}\left(\frac{1}{n^2}\right)$. Therefore, using (the right-hand part of) (6.14), we obtain

$$\frac{1}{1+\left(\dfrac{2}{n}+\dfrac{1}{4n^2}\right)} = 1-\frac{2}{n}-\frac{1}{4n^2}+\mathcal{O}\left(\frac{2}{n}+\frac{1}{4n^2}\right)^2 = 1-\frac{2}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)+\mathcal{O}\left(\frac{1}{n^2}\right)$$
$$= 1-\frac{2}{n}+\mathcal{O}\left(\frac{1}{n^2}\right),$$

since $\mathcal{O}(2/n^2) = \mathcal{O}(1/n^2)$.

Here we can see the very "big" advantage of using the big $\mathcal{O}$ notation: it hides a lot of complicated junk information. For example, the left-hand side of the equation is exactly equal to (see the left-hand part of (6.14))

$$\frac{1}{1+\left(\dfrac{2}{n}+\dfrac{1}{4n^2}\right)} = 1-\frac{2}{n}+\left[-\frac{1}{4n^2}+\frac{\left(\frac{2}{n}+\frac{1}{4n^2}\right)^2}{1+\frac{2}{n}+\frac{1}{4n^2}}\right],$$

so the big $\mathcal{O}$ notation allows us to summarize the complicated material on the right as the very simple $\mathcal{O}\big(\frac{1}{n^2}\big)$.

Example 6.22. Consider our "mystery" series

$$1+\sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots(2n)\,(2n+1)}$$

already considered in (6.9). We saw that the ratio and root tests failed for this series; however, it turns out that Raabe's test works. To see this, let a_n denote the n-th term in the "mystery" series. Then from (6.10), we see that

$$\frac{a_n}{a_{n+1}} = \frac{1+\dfrac{5}{2n}+\dfrac{3}{2n^2}}{1+\dfrac{2}{n}+\dfrac{1}{4n^2}} = \left(1+\frac{5}{2n}+\frac{3}{2n^2}\right)\left(1-\frac{2}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)$$
$$= \left(1+\frac{5}{2n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)\left(1-\frac{2}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right).$$

Multiplying out the right-hand side, using the properties of big $\mathcal{O}$, we get

$$\frac{a_n}{a_{n+1}} = 1+\frac{5}{2n}-\frac{2}{n}+\mathcal{O}\left(\frac{1}{n^2}\right) = 1+\frac{1}{2n}+\mathcal{O}\left(\frac{1}{n^2}\right).$$

Hence,

$$n\left(\frac{a_n}{a_{n+1}}-1\right) = \frac{1}{2}+\mathcal{O}\left(\frac{1}{n}\right) \quad\Longrightarrow\quad \lim n\left(\frac{a_n}{a_{n+1}}-1\right) = \frac{1}{2} > 0,$$

so by Raabe's test, the "mystery" sum converges.[4]

[4] It turns out that the "mystery" sum equals $\pi/2$; see [**136**] for a proof.

6.3.3. De Morgan and Bertrand’s test. We next study a test due to Augustus De Morgan (1806–1871) and Joseph Bertrand (1822–1900). For this test, we let $b_n = 1/n\log n$ in Kummer’s test.

Theorem 6.16 (**De Morgan and Bertrand’s test**). *Let $\{a_n\}$ be a sequence of positive numbers and define α_n by the equation*

$$\frac{a_n}{a_{n+1}} = 1 + \frac{1}{n} + \frac{\alpha_n}{n\log n}.$$

Then $\sum a_n$ converges or diverges according as $\liminf \alpha_n > 1$ *or* $\limsup \alpha_n < 1$.

Proof. If we let $b_n = 1/n\log n$ in Kummer’s test, then

$$\begin{aligned}\kappa_n = \frac{1}{b_n}\frac{a_n}{a_{n+1}} - \frac{1}{b_{n+1}} &= n\log n\left(1 + \frac{1}{n} + \frac{\alpha_n}{n\log n}\right) - (n+1)\log(n+1)\\ &= \alpha_n + (n+1)\Big[\log n - \log(n+1)\Big].\end{aligned}$$

Since

$$(n+1)\Big[\log n - \log(n+1)\Big] = \log\left(1 - \frac{1}{n+1}\right)^{n+1} \to \log e^{-1} = -1,$$

we have

$$\liminf \kappa_n = \liminf \alpha_n - 1 \quad \text{and} \quad \limsup \kappa_n = \limsup \alpha_n - 1.$$

Invoking Kummer’s test now completes the proof. □

6.3.4. Gauss’s test. Finally, to end our potpourri of tests, we conclude with Gauss’ test:

Theorem 6.17 (**Gauss’ test**). *Let $\{a_n\}$ be a sequence of positive numbers and suppose that we can write*

$$\frac{a_n}{a_{n+1}} = 1 + \frac{\xi}{n} + \mathcal{O}\left(\frac{1}{n^p}\right),$$

where ξ is a constant and $p > 1$. Then $\sum a_n$ converges or diverges according as $\xi \le 1$ or $\xi > 1$.

Proof. The hypotheses imply that

$$n\left(\frac{a_n}{a_{n+1}} - 1\right) = \xi + n\mathcal{O}\left(\frac{1}{n^p}\right) = \xi + \mathcal{O}\left(\frac{1}{n^{p-1}}\right) \to \xi$$

as $n \to \infty$, where we used that $p-1 > 0$. Thus, Raabe’s test shows that series $\sum a_n$ converges for $\xi > 1$ and diverges for $\xi < 1$. For the case $\xi = 1$, let $\frac{a_n}{a_{n+1}} = 1 + \frac{1}{n} + f_n$ where $f_n = \mathcal{O}\left(\frac{1}{n^p}\right)$. Then we can write

$$\frac{a_n}{a_{n+1}} = 1 + \frac{1}{n} + f_n = 1 + \frac{1}{n} + \frac{\alpha_n}{n\log n},$$

where $\alpha_n = f_n\, n\log n$. If we let $p = 1+\delta$, where $\delta > 0$, then we know that $\frac{\log n}{n^\delta} \to 0$ as $n \to \infty$ by Problem 8 in Exercises 4.6, so

$$\alpha_n = f_n\, n\log n = \mathcal{O}\left(\frac{1}{n^{1+\delta}}\right) n\log n = \mathcal{O}\left(\frac{\log n}{n^\delta}\right) \quad \Longrightarrow \quad \lim \alpha_n = 0.$$

Thus, De Morgan and Bertrand’s test shows that the series $\sum a_n$ diverges. □

Example 6.23. Gauss' test originated with Gauss' study of the hypergeometric series:

$$1+\frac{\alpha\cdot\beta}{1\cdot\gamma}+\frac{\alpha(\alpha-1)\cdot\beta(\beta-1)}{2!\cdot\gamma(\gamma+1)}+\frac{\alpha(\alpha-1)(\alpha-2)\cdot\beta(\beta-1)(\beta-2)}{3!\cdot\gamma(\gamma+1)(\gamma+2)}+\cdots,$$

where α,β,γ are positive real numbers. We can write this as $\sum a_n$ where

$$a_n=\frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)\cdot\beta(\beta-1)(\beta-2)\cdot(\beta-n+1)}{n!\cdot\gamma(\gamma+1)(\gamma+2)\cdots(\gamma+n-1)}.$$

Hence, for $n\geq 1$ we have

$$\frac{a_n}{a_{n+1}}=\frac{(n+1)(\gamma+n)}{(\alpha+n)(\beta+n)}=\frac{n^2+(\gamma+1)n+\gamma}{n^2+(\alpha+\beta)n+\alpha\beta}=\frac{1+\dfrac{\gamma+1}{n}+\dfrac{\gamma}{n^2}}{1+\dfrac{\alpha+\beta}{n}+\dfrac{\alpha\beta}{n^2}}.$$

Using the handy formula from (6.14),

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

we see that (after some algebra)

$$\begin{aligned}\frac{a_n}{a_{n+1}}&=\left(1+\frac{\gamma+1}{n}+\frac{\gamma}{n^2}\right)\left[1-\frac{\alpha+\beta}{n}-\frac{\alpha\beta}{n^2}+\mathcal{O}\left(\frac{1}{n^2}\right)\right]\\&=1+\frac{\gamma+1-\alpha-\beta}{n}+\mathcal{O}\left(\frac{1}{n^2}\right).\end{aligned}$$

Thus, the hypergeometric series converges if $\gamma>\alpha+\beta$ and diverges if $\gamma\leq\alpha+\beta$.

Exercises 6.3.

1. Determine whether or not the following series converge.

$$(a)\ \sum_{n=1}^{\infty}\frac{1\cdot3\cdot5\cdots(2n-1)}{2^n(n+1)!}\quad,\quad(b)\ \sum_{n=1}^{\infty}\frac{3\cdot6\cdot9\cdots(3n)}{7\cdot10\cdot13\cdots(3n+4)},$$

$$(c)\ \sum_{n=1}^{\infty}\frac{1\cdot3\cdot5\cdots(2n-1)}{2\cdot4\cdot6\cdots(2n)}\quad,\quad(d)\ \sum_{n=1}^{\infty}\frac{2\cdot4\cdot6\cdots(2n+2)}{1\cdot3\cdot5\cdots(2n-1)(2n)}.$$

For $\alpha,\beta\neq 0,-1,-2,\ldots,$

$$(e)\ \sum_{n=1}^{\infty}\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{n!}\quad,\quad(f)\ \sum_{n=1}^{\infty}\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{\beta(\beta+1)(\beta+2)\cdots(\beta+n-1)}.$$

If $\alpha,\beta,\gamma,\kappa,\lambda\neq 0,-1,-2,\ldots,$ then prove that the following monster

$$(g)\ \sum_{n=1}^{\infty}\frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots(\beta+n-1)\gamma(\gamma+1)\cdots(\gamma+n-1)}{n!\,\kappa(\kappa+1)\cdots(\kappa+n-1)\lambda(\lambda+1)\cdots(\lambda+n-1)}$$

converges for $\kappa+\lambda-\alpha-\beta-\gamma>0$.

2. Using Raabe's test, prove that $\sum 1/n^p$ converges for $p>1$ and diverges for $p<1$.
3. (**Logarithmic test**) We prove a useful test called the **logarithmic test**: If $\sum a_n$ is a series of positive terms, then this series converges or diverges according as

$$\liminf\left(n\log\frac{a_n}{a_{n+1}}\right)>1\quad\text{or}\quad\limsup\left(n\log\frac{a_n}{a_{n+1}}\right)<1.$$

To prove this, proceed as follows.

(i) Suppose first that $\liminf \left(n \log \frac{a_n}{a_{n+1}}\right) > 1$. Show that there is an $a > 1$ and an N such that

$$n > N \quad \Longrightarrow \quad a < n \log \frac{a_n}{a_{n+1}} \quad \Longrightarrow \quad \frac{a_{n+1}}{a_n} < e^{-a/n}.$$

(ii) Using $\left(1 + \frac{1}{n}\right)^n < e$ from (3.28), the p-test, and the limit comparison test (see Problem 7 in Exercises 3.6), prove that $\sum a_n$ converges.

(iii) Similarly, prove that if $\limsup \left(n \log \frac{a_n}{a_{n+1}}\right) < 1$, then $\sum a_n$ diverges.

(iv) Using the logarithmic test, determine the convergence/diverence of

$$\sum_{n=1}^{\infty} \frac{n!}{n^n} \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{n^n}{n!}.$$

6.4. Some pretty powerful properties of power series

The title of this section speaks for itself. As stated already, we focus on power series of a complex variable z, but all the results stated in this section have corresponding statements for power series of a real variable x.

6.4.1. Continuity and the exponential function (again).

We first prove that power series are always continuous (within their radius of convergence).

Lemma 6.18. *If $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R, then $\sum_{n=1}^{\infty} n\, a_n z^{n-1}$ also has radius of convergence R.*

Proof. (See Problem 3 for another proof of this lemma using properties of $\limsup$.) For $z \neq 0$, $\sum_{n=1}^{\infty} n\, a_n z^{n-1}$ converges if and only if $z \cdot \sum_{n=1}^{\infty} n\, a_n z^{n-1} = \sum_{n=1}^{\infty} n\, a_n z^n$ converges, so we just have to show that $\sum_{n=1}^{\infty} n\, a_n z^n$ has radius of convergence R. Since $|a_n| \leq n|a_n|$, by comparison, if $\sum_{n=1}^{\infty} n\, |a_n|\, |z|^n$ converges, then $\sum_{n=1}^{\infty} |a_n|\, |z|^n$ also converges, so the radius of convergence of the series $\sum_{n=1}^{\infty} n\, a_n z^n$ can't be larger than R. To prove that the radius of convergence is at least R, fix z with $|z| < R$; we need to prove that $\sum_{n=1}^{\infty} n\, |a_n|\, |z|^n$ converges. To this end, fix ρ with $|z| < \rho < R$ and note that $\sum_{n=1}^{\infty} n\, (|z|/\rho)^n$ converges, by e.g. the root test:

$$\lim \left| n \left(\frac{|z|}{\rho}\right)^n \right|^{1/n} = \lim n^{1/n} \cdot \frac{|z|}{\rho} = \frac{|z|}{\rho} < 1.$$

Since $\sum_{n=1}^{\infty} |a_n| \rho^n$ converges (because $\rho < R$, the radius of convergence of the series $\sum_{n=0}^{\infty} a_n z^n$), by the n-th term test, $|a_n|\rho^n \to 0$ as $n \to \infty$. In particular, $|a_n|\rho^n \leq M$ for some constant M, hence

$$n\, |a_n|\, |z|^n = n\, |a_n|\, \rho^n \cdot \left(\frac{|z|}{\rho}\right)^n \leq M \cdot n \left(\frac{|z|}{\rho}\right)^n.$$

Since $M \sum n\, (|z|/\rho)^n$ converges, by the comparison test, it follows that $\sum n\, |a_n|\, |z|^n$ also converges. This completes our proof. □

Theorem 6.19 (**Continuity theorem for power series**). *A power series is continuous within its radius of convergence.*

Proof. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ have radius of convergence R; we need to show that $f(z)$ is continuous at each point $c \in \mathbb{C}$ with $|c| < R$. So, let us fix such a c. Since

$$z^n - c^n = (z - c)\, q_n(z), \quad \text{where} \quad q_n(z) = z^{n-1} + z^{n-2}\, c + \cdots + z\, c^{n-2} + c^{n-1},$$

which is proved by multiplying out $(z-c)\, q_n(z)$, we can write

$$f(z) - f(c) = \sum_{n=1}^{\infty} a_n (z^n - c^n) = (z-c) \sum_{n=0}^{\infty} a_n q_n(z).$$

To make the sum $\sum_{n=0}^{\infty} a_n q_n(z)$ small in absolute value we proceed as follows. Fix r such that $|c| < r < R$. Then for $|z-c| < r - |c|$, we have

$$|z| \le |z-c| + |c| < r - |c| + |c| = r.$$

Thus, as $|c| < r$, for $|z-c| < r-|c|$ we see that

$$|q_n(z)| \le \underbrace{r^{n-1} + r^{n-2}\, r + \cdots + r\, r^{n-2} + r^{n-1}}_{n \text{ terms}} = n r^{n-1}.$$

By our lemma, $\sum_{n=1}^{\infty} n\, |a_n|\, r^{n-1}$ converges, so if $C := \sum_{n=1}^{\infty} n\, |a_n|\, r^{n-1}$, then

$$|f(z) - f(c)| \le |z-c| \sum_{n=1}^{\infty} |a_n|\, |q_n(z)| \le |z-c| \sum_{n=1}^{\infty} |a_n|\, n r^{n-1} = C|z-c|,$$

which implies that $\lim_{z \to c} f(z) = f(c)$; that is, f is continuous at $z = c$. □

6.4.2. Abel's limit theorem. Abel's limit theorem has to do with the following question. Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence R; this implies, in particular, that $f(x)$ is defined for all $-R < x < R$ and, by Theorem 6.19, is continuous on the interval $(-R, R)$. Let us suppose that $f(R) = \sum_{n=0}^{\infty} a_n R^n$ converges. In particular, $f(x)$ is defined for all $-R < x \le R$. Question: Is f continuous on the interval $(-R, R]$, that is, is it true that

$$\lim_{x \to R-} f(x) = f(R)? \tag{6.16}$$

The answer to this question is "yes" and it follows from the following more general theorem due to Neils Abel; however, Abel's theorem is mostly used for the real variable case $\lim_{x \to R-} f(x) = f(R)$ that we just described.

Theorem 6.20 (**Abel's limit theorem**). *Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ have radius of convergence R and let $z_0 \in \mathbb{C}$ with $|z_0| = R$ where the series $f(z_0) = \sum_{n=0}^{\infty} a_n z_0^n$ converges. Then*

$$\lim_{z \to z_0} f(z) = f(z_0)$$

where the limit on the left is taken in such a way that $|z| < R$ and that the ratio $\frac{|z_0 - z|}{R - |z|}$ remains bounded by a fixed constant.

Proof. By considering the limit of the function $g(z) = f(z_0 z) - f(z_0)$ as $z \to 1$ in such a way that $|z| < 1$ and that the ratio $|1-z|/(1-|z|)$ remains bounded by a fixed constant, we may henceforth assume that $z_0 = 1$ and that $f(z_0) = 0$ (the diligent student will check the details of this statement). With these assumptions, if we put $s_n = a_0 + a_1 + \cdots + a_n$, then $0 = f(1) = \sum_{n=0}^{\infty} a_n = \lim s_n$. Now observe that $a_n = s_n - s_{n-1}$, so

$$\begin{aligned}\sum_{k=0}^{n} a_k z^k &= a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n \\ &= s_0 + (s_1 - s_0)z + (s_2 - s_1)z^2 + \cdots + (s_n - s_{n-1})z^n \\ &= s_0(1-z) + s_1(z - z^2) + \cdots + s_{n-1}(z^{n-1} - z^n) + s_n z^n \\ &= s_0(1-z) + s_1(1-z)z + \cdots + s_{n-1}(1-z)z^{n-1} + s_n z^n \\ &= (1-z)\big(s_0 + s_1 z + \cdots + s_{n-1}z^{n-1}\big) + s_n z^n.\end{aligned}$$

Thus, $\sum_{k=0}^{n} a_k z^k = (1-z)\sum_{k=0}^{n} s_k z^k + s_n z^n$. Since $s_n \to 0$ and $|z| < 1$ it follows that $s_n z^n \to 0$. Therefore, taking $n \to \infty$, we obtain

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = (1-z)\sum_{n=0}^{\infty} s_n z^n,$$

which implies that

$$|f(z)| \leq |1-z| \sum_{n=0}^{\infty} |s_n|\,|z|^n.$$

Let us now take $z \to 1$ in such a way that $|z| < 1$ and $|1-z|/(1-|z|) < C$ where $C > 0$. Let $\varepsilon > 0$ be given and, since $s_n \to 0$, we can choose an integer N such that $n > N \Longrightarrow |s_n| < \varepsilon/(2C)$. Define $K := \sum_{n=0}^{N} |s_n|$. Then we can write

$$\begin{aligned}|f(z)| &\leq |1-z| \sum_{n=0}^{N} |s_n|\,|z|^n + |1-z| \sum_{n=N}^{\infty} |s_n|\,|z|^n \\ &< |1-z| \sum_{n=0}^{N} |s_n| \cdot 1^n + |1-z| \sum_{n=N}^{\infty} \frac{\varepsilon}{2C}\,|z|^n \\ &= K|1-z| + \frac{\varepsilon}{2C}|1-z| \sum_{n=0}^{\infty} |z|^n \\ &= K|1-z| + \frac{\varepsilon}{2C}\frac{|1-z|}{1-|z|} < K|1-z| + \frac{\varepsilon}{2}.\end{aligned}$$

Thus, with $\delta := \varepsilon/(2K)$, we have

$$|z-1| < \delta \;\text{ with }\; |z| < 1 \;\text{ and }\; \frac{|1-z|}{1-|z|} < C \quad \Longrightarrow \quad |f(z)| < \varepsilon.$$

This completes our proof. □

Notice that for $z = x$ with $0 < x < R$, we have

$$\frac{|R-z|}{R-|z|} = \frac{|R-x|}{R-|x|} = \frac{R-x}{R-x} = 1$$

which, in particular, is bounded by 1, so (6.16) holds under the assumptions stated. Once we prove this result at $x = R$, we can prove a similar result at $x = -R$: If $f(x) = \sum_{n=0}^{\infty} a_n x^n$ has radius of convergence R and $f(-R) = \sum_{n=0}^{\infty} a_n (-R)^n$ converges, then

$$\lim_{x \to -R+} f(x) = f(-R).$$

To prove this, consider the function $g(x) = f(-x)$, then apply (6.16) to g.

6.4.3. The identity theorem.

6.4.3. The identity theorem. The identity theorem is perhaps one of the most useful properties of power series. The identity theorem says, very roughly, that if two power series are identical at "sufficiently many" points, then in fact, the power series are identical everywhere!

Theorem 6.21 (**Identity theorem**). *Let $f(z) = \sum a_n z^n$ and $g(z) = \sum b_n z^n$ have positive radii of convergence and suppose that $f(c_k) = g(c_k)$ for some nonzero sequence $c_k \to 0$. Then the power series $f(z)$ and $g(z)$ must be identical; that is $a_n = b_n$ for every $n = 0, 1, 2, 3, \ldots$.*

Proof. We begin by proving that for each $m = 0, 1, 2, \ldots$, the series

$$f_m(z) := \sum_{n=m}^{\infty} a_n z^{n-m} = a_m + a_{m+1} z + a_{m+2} z^2 + a_{m+3} z^3 + \cdots$$

has the same radius of convergence as f. Indeed, since we can write

$$f_m(z) = z^{-m} \sum_{n=m}^{\infty} a_n z^n$$

for $z \neq 0$, the power series $f_m(z)$ converges if and only if $\sum_{n=m}^{\infty} a_n z^n$ converges, which in turn converges if and only if $f(z)$ converges. It follows that $f_m(z)$ and $f(z)$ have the same radius of convergence; in particular, by the continuity theorem for power series, $f_m(z)$ is continuous at 0. Similarly, for each $m = 0, 1, 2, \ldots$, $g_m(z) := \sum_{n=m}^{\infty} b_n z^{n-m}$ has the same radius of convergence as $g(z)$; in particular, $g_m(z)$ is continuous at 0. These continuity facts concerning f_m and g_m are the important facts that will be used below.

Now to our proof. We are given that

$$a_0 + a_1 c_k + a_2 c_k^2 + \cdots = b_0 + b_1 c_k + b_2 c_k^2 + \cdots \quad \text{that is, } f(c_k) = g(c_k) \tag{6.17}$$

for all k. In particular, taking $k \to \infty$ in the equality $f(c_k) = g(c_k)$, using that $c_k \to 0$ and that f and g are continuous at 0, we obtain $f(0) = g(0)$, or $a_0 = b_0$. Cancelling $a_0 = b_0$ and dividing by $c_k \neq 0$ in (6.17), we obtain

$$a_1 + a_2 c_k + a_3 c_k^2 + \cdots = b_1 + b_2 c_k + b_3 c_k^2 + \cdots \quad \text{that is, } f_1(c_k) = g_1(c_k) \tag{6.18}$$

for all k. Taking $k \to \infty$ and using that $c_k \to 0$ and that f_1 and g_1 are continuous at 0, we obtain $f_1(0) = g_1(0)$, or $a_1 = b_1$. Cancelling $a_1 = b_1$ and dividing by $c_k \neq 0$ in (6.18), we obtain

$$a_2 + a_3 c_k + a_4 c_k^2 + \cdots = b_2 + b_3 c_k + b_4 c_k^2 + \cdots \quad \text{that is, } f_2(c_k) = g_2(c_k) \tag{6.19}$$

for all k. Taking $k \to \infty$, using that $c_k \to 0$ and that f_2 and g_2 are continuous at 0, we obtain $f_2(0) = g_2(0)$, or $a_2 = b_2$. Continuing by induction we get $a_n = b_n$ for all $n = 0, 1, 2, \ldots$, which is exactly what we wanted to prove. □

Corollary 6.22. *If $f(z) = \sum a_n z^n$ and $g(z) = \sum b_n z^n$ have positive radii of convergence and $f(x) = g(x)$ for all $x \in \mathbb{R}$ with $|x| < \varepsilon$ for some $\varepsilon > 0$, then $a_n = b_n$ for every n; in other words, f and g are actually the same power series.*

Proof. To prove this, observe that since $f(x) = g(x)$ for all $x \in \mathbb{R}$ such that $|x| < \varepsilon$, then $f(c_k) = g(c_k)$ for all k sufficiently large where $c_k = 1/k$; the identity theorem now implies $a_n = b_n$ for every n. □

Using the identity theorem we can deduce certain properties of series.

Example 6.24. Suppose that $f(z) = \sum a_n z^n$ is an **odd function** in the sense that $f(-z) = -f(z)$ for all z within its radius of convergence. In terms of power series, the identity $f(-z) = -f(z)$ is

$$\sum a_n (-1)^n z^n = \sum -a_n z^n.$$

By the identity theorem, we must have $(-1)^n a_n = -a_n$ for each n. Thus, for n even we must have $a_n = -a_n$ or $a_n = 0$, and for n odd, we must have $-a_n = -a_n$, a tautology. In conclusion, we see that f is odd if and only if all coefficients of even powers vanish:

$$f(z) = \sum_{n=0}^{\infty} a_{2n+1} z^{2n+1};$$

that is, f is odd if and only if f has only odd powers in its series expansion.

Exercises 6.4.

1. Prove that $f(z) = \sum a_n z^n$ is an **even function** in the sense that $f(-z) = f(z)$ for all z within its radius of convergence if and only if f has only even powers in its expansion, that is, f takes the form $f(z) = \sum_{n=0}^{\infty} a_{2n} z^{2n}$.
2. Recall that the binomial coefficient is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for $0 \le k \le n$. Prove the highly nonobvious result:
$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j} \binom{n}{k-j}.$$
Suggestion: Apply the binomial formula to $(1+z)^{m+n}$, which equals $(1+z)^m \cdot (1+z)^n$. Prove that
$$\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2.$$
3. Prove that $\sum_{n=1}^{\infty} n\,|a_n|\,r^n$ converges, where the notation is as in the proof of Theorem 6.19, using the root test. You will need Problem 9 in Exercises 6.2.
4. (**Abel summability**) We say that a series $\sum a_n$ is **Abel summable** to L if the power series $f(x) := \sum a_n x^n$ is defined for all $x \in [0,1)$ and $\lim_{x \to 1-} f(x) = L$.
   (a) Prove that if $\sum a_n$ converges to $L \in \mathbb{C}$, then $\sum a_n$ is also Abel summable to L.
   (b) Derive the following amazing formulas (properly interpreted!):
$$\boxed{\begin{aligned} &1 - 1 + 1 - 1 + 1 - 1 + - \cdots =_a \frac{1}{2}, \\ &1 + 2 - 3 + 4 - 5 + 6 - 7 + - \cdots =_a \frac{1}{4}, \end{aligned}}$$
   where $=_a$ mean "is Abel summable to". You will need Problem 6 in Exercises 3.5.
5. In this problem we continue our fascinating study of Abel summability. Let $a_0, a_1, a_2, \ldots$ be a positive nonincreasing sequence tending to zero (in particular, $\sum (-1)^{n-1} a_n$ converges by the alternating series test). Define $b_n := a_0 + a_1 + \cdots + a_n$. We shall prove the neat formula
$$b_0 - b_1 + b_2 - b_3 + b_4 - b_5 + - \cdots =_a \frac{1}{2} \sum_{n=0}^{\infty} (-1)^n a_n.$$
   (i) Let $f(x) = \sum_{n=0}^{\infty} (-1)^n b_n\, x^n$. Prove that f has radius of convergence 1. Suggestion: Use the ratio test.

(ii) Let

$$\begin{aligned} f_n(x) &= \sum_{k=0}^{n} (-1)^k b_k \, x^k \\ &= a_0 - (a_0 + a_1)x + (a_0 + a_1 + a_2)x^2 - \cdots + (-1)^n (a_0 + a_1 + \cdots + a_n)x^n \end{aligned}$$

be the n-th partial sum of $f(x)$. Prove that

$$f_n(x) = \frac{1}{1+x}\big(a_0 - a_1 x + a_2 x^2 - a_3 x^3 + \cdots + (-1)^n a_n x^n\big) \\ + (-1)^n \frac{x^{n+1}}{1+x}\big(a_0 + a_2 + a_3 + \cdots + a_n\big).$$

(iii) Prove that[5]

$$f(x) = \frac{1}{1+x} \sum_{n=0}^{\infty} (-1)^n a_n x^n.$$

Finally, from this formula prove the desired result.

(iv) Establish the remarkable formula

$$\boxed{1 - \Big(1 + \frac{1}{2}\Big) + \Big(1 + \frac{1}{2} + \frac{1}{3}\Big) - \Big(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\Big) + - \cdots =_a \frac{1}{2}\log 2.}$$

6. Suppose that $f(z) = \sum a_n z^n$ has radius of convergence 1, where $\sum a_n$ is a divergent series of positive real numbers. Prove that $\lim_{x \to 1-} f(x) = +\infty$.

6.5. Double sequences, double series, and a ζ-function identity

After studying single integrals in elementary calculus, you probably took a course where you studied “double integrals”. In a similar way, now that we have a thorough background in “single infinite series,” we now move to the topic of “double infinite series”. The main result of this section is Cauchy’s double series theorem — Theorem 6.26, which we’ll use quite often in the sequel. If you did Problem 9 in Exercises 3.7 you tasted a bit of Cauchy’s theorem in its relation to Tannery’s theorem (however, we won’t assume Tannery’s theorem for this section). The books [**144**, Ch. 3] and [**41**, Ch. 5] have lots of material on double sequences and series.

6.5.1. Double sequences and series and Pringsheim’s theorem. We begin by studying double sequences. Recall that a complex sequence is really just a function $s : \mathbb{N} \to \mathbb{C}$ where we usually denote $s(n)$ by s_n. By analogy, we define a **double sequence** of complex numbers as a function $s : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{C}$. We usually denote $s(m,n)$ by s_{mn} and the corresponding double sequence by $\{s_{mn}\}$.

Example 6.25. For $m, n \in \mathbb{N}$,

$$s_{mn} = \frac{m \cdot n}{(m+n)^2}$$

defines a double sequence $\{s_{mn}\}$.

Whenever we talk about sequences, the idea of convergence is bound to follow. Let $\{s_{mn}\}$ be a double sequence of complex numbers. We say that the double sequence $\{s_{mn}\}$ **converges** if there is a complex number L having the property that given any $\varepsilon > 0$ there is a real number N such that

$$m, n > N \quad \Longrightarrow \quad |L - s_{mn}| < \varepsilon,$$

[5] In the next section, we’ll learn how to prove this identity in a much quicker way using the technologically advanced Cauchy’s double series theorem.

in which case we write $L = \lim s_{mn}$.

Care has to be taken when dealing with double sequences because sometimes sequences that look convergent are actually not.

Example 6.26. The nice looking double sequence $s_{mn} = mn/(m+n)^2$ does not converge. To see this, observe that if $m = n$, then

$$s_{mn} = \frac{n \cdot n}{(n+n)^2} = \frac{n^2}{4n^2} = \frac{1}{4}.$$

However, if $m = 2n$, then

$$s_{mn} = \frac{2n \cdot n}{(2n+n)^2} = \frac{2n^2}{9n^2} = \frac{2}{9}.$$

Therefore it is impossible for s_{mn} to approach any single number no matter how large we take m, n.

Given a double sequence $\{s_{mn}\}$ it is convenient to look at the **iterated limits**:

$$\lim_{m\to\infty} \lim_{n\to\infty} s_{mn} \qquad \text{and} \qquad \lim_{n\to\infty} \lim_{m\to\infty} s_{mn}. \tag{6.20}$$

For $\lim_{m\to\infty} \lim_{n\to\infty} s_{mn}$ on the left, we mean to first take $n \to \infty$ and second to take $m \to \infty$, reversing the order for $\lim_{n\to\infty} \lim_{m\to\infty} s_{mn}$. In general, the iterated limits (6.20) may have no relationship!

Example 6.27. Consider the double sequence $s_{mn} = mn/(m+n^2)$. We have

$$\lim_{n\to\infty} s_{mn} = \lim_{n\to\infty} \frac{mn}{m+n^2} = 0 \quad \Longrightarrow \quad \lim_{m\to\infty} \lim_{n\to\infty} s_{mn} = \lim_{m\to\infty} 0 = 0.$$

On the other hand,

$$\lim_{m\to\infty} s_{mn} = \lim_{m\to\infty} \frac{mn}{m+n^2} = n \quad \Longrightarrow \quad \lim_{n\to\infty} \lim_{m\to\infty} s_{mn} = \lim_{n\to\infty} n = \infty.$$

Here are a couple questions:

(I) If both iterated limits (6.20) exist and are equal, say to a number L, is it true that the regular double limit $\lim s_{mn}$ exists and $\lim s_{mn} = L$?

(II) If $L = \lim s_{mn}$ exists, is it true that both iterated limits (6.20) exist and are equal to L:

$$L = \lim_{m\to\infty} \lim_{n\to\infty} s_{mn} = \lim_{n\to\infty} \lim_{m\to\infty} s_{mn}? \tag{6.21}$$

It may shock you, but the answer to both of these questions is "no".

Example 6.28. For a counter example to Question I, consider our first example $s_{mn} = mn/(m+n)^2$. We know that $\lim s_{mn}$ does not exist, but observe that

$$\lim_{n\to\infty} s_{mn} = \lim_{n\to\infty} \frac{mn}{(m+n)^2} = 0 \quad \Longrightarrow \quad \lim_{m\to\infty} \lim_{n\to\infty} s_{mn} = \lim_{m\to\infty} 0 = 0.$$

and

$$\lim_{m\to\infty} s_{mn} = \lim_{m\to\infty} \frac{mn}{(m+n)^2} = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} \lim_{m\to\infty} s_{mn} = \lim_{n\to\infty} 0 = 0,$$

so both iterated limits converge. For a counter example to Question II, see limit (d) in Problem 1.

However, if a double sequence converges and both iterated limits exists, then they all must equal the same number. This is the content of the following theorem, named after Alfred Pringsheim (1850–1941) (cf. [**41**, p. 79]).

Theorem 6.23 (**Pringsheim's theorem for sequences**). *If $\{s_{mn}\}$ converges and for each m, $\lim_{n\to\infty} s_{mn}$ exists and for each n, $\lim_{m\to\infty} s_{mn}$ exists, then both iterated limits exist and the equality* (6.21) *holds.*

Proof. Let $\varepsilon > 0$. Then there is an N such that for all $m, n > N$, we have $|L - s_{mn}| < \varepsilon/2$. Taking $n \to \infty$, we get, for $m > N$, $|L - \lim_{n\to\infty} s_{mn}| \leq \varepsilon/2$. Hence,

$$m > N \quad \Longrightarrow \quad \big|L - \lim_{n\to\infty} s_{mn}\big| < \varepsilon.$$

This means that $\lim_{m\to\infty}(\lim_{n\to\infty} s_{mn}) = L$. A similar argument establishes the equality with the limits of m and n reversed. □

Recall that if $\{a_n\}$ is a sequence of complex numbers, then we say that $\sum a_n$ converges if the sequence $\{s_n\}$ converges, where $s_n := \sum_{k=1}^n a_k$. By analogy, we define a double series of complex numbers as follows. Let $\{a_{mn}\}$ be a double sequence of complex numbers and let

$$s_{mn} := \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij},$$

called the m, n**-th partial sum** of $\sum a_{mn}$. We say that the double series $\sum a_{mn}$ **converges** if the double sequence $\{s_{mn}\}$ of partial sums converges. If $\sum a_{mn}$ exists, we can ask whether or not

$$\sum a_{mn} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} \ ? \tag{6.22}$$

Here, with $s_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}$, the **iterated series** on the right are defined as

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} := \lim_{m\to\infty} \lim_{n\to\infty} s_{mn} \quad \text{and} \quad \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} := \lim_{n\to\infty} \lim_{m\to\infty} s_{mn}.$$

Thus, (6.22) is just the equality (6.21) with $s = \sum a_{mn}$. Hence, Pringsheim's theorem for sequences immediately implies the following.

Theorem 6.24 (**Pringsheim's theorem for series**). *If a double series $\sum a_{mn}$ converges and for each m, $\sum_{n=1}^{\infty} a_{mn}$ converges and for each n, $\sum_{m=1}^{\infty} a_{mn}$ converges, then both iterated series converge and the equality* (6.22) *holds.*

We can "visualize" the iterated sums in (6.22) as follows. First, we arrange the a_{mn}'s in an infinite array as shown in Figure 6.4. Then for fixed $m \in \mathbb{N}$, the sum $\sum_{n=1}^{\infty} a_{mn}$ is summing all the numbers in the m-th row shown in the left picture in Figure 6.4. For example, if $m = 1$, then $\sum_{n=1}^{\infty} a_{1n}$ is summing all the numbers in the first row shown in the left picture in Figure 6.4. The summation $\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn}$ is summing over all the rows (that have already been summed). Similarly, $\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn}$ is summing over all the columns. In Subsection 6.5.3 we shall study the most useful theorem on iterated sums, Cauchy's double series theorem, which states that (6.22) always holds for absolutely convergent series. Here, a double series $\sum a_{mn}$ is said to **converge absolutely** if the double series of absolute values $\sum |a_{mn}|$ converges. However, before presenting Cauchy's theorem, we first generalize summing by rows and columns to "summing by curves".

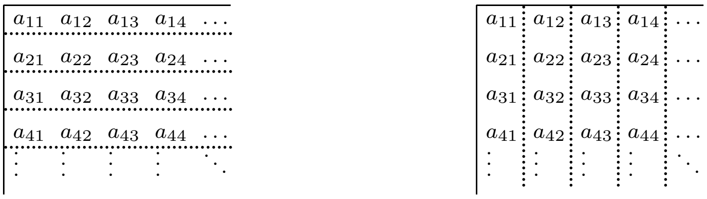

a_{11} a_{12} a_{13} a_{14} …
a_{21} a_{22} a_{23} a_{24} …
a_{31} a_{32} a_{33} a_{34} …
a_{41} a_{42} a_{43} a_{44} …
a_{11} a_{12} a_{13} a_{14} …
a_{21} a_{22} a_{23} a_{24} …
a_{31} a_{32} a_{33} a_{34} …
a_{41} a_{42} a_{43} a_{44} …

FIGURE 6.4. In the first array we are "summing by rows" and in the second array we are "summing by columns".

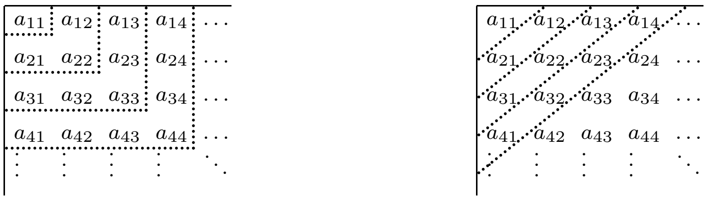

a_{11} a_{12} a_{13} a_{14} …
a_{21} a_{22} a_{23} a_{24} …
a_{31} a_{32} a_{33} a_{34} …
a_{41} a_{42} a_{43} a_{44} …
a_{11} a_{12} a_{13} a_{14} …
a_{21} a_{22} a_{23} a_{24} …
a_{31} a_{32} a_{33} a_{34} …
a_{41} a_{42} a_{43} a_{44} …

FIGURE 6.5. "Summing by squares" and "summing by triangles".

6.5.2. "Summing by curves". Before presenting the "sum by curves theorem" (Theorem 6.25 below) it might be helpful to give a couple examples of this theorem to help in understanding what it says. Let $\sum a_{mn}$ be a double series.

Example 6.29. Let

$$S_k = \{(m,n)\,;\, 1 \le m \le k\ ,\ 1 \le n \le k\},$$

which represents a $k \times k$ square of numbers; see the left-hand picture in Figure 6.5 for 1×1, 2×2, 3×3, and 4×4 examples. We denote by $\sum_{(m,n)\in S_k} a_{mn}$ the sum of those a_{mn}'s within the $k \times k$ square S_k. Explicitly,

$$\sum_{(m,n)\in S_k} a_{mn} = \sum_{m=1}^{k}\sum_{n=1}^{k} a_{mn}.$$

It is natural to refer to the limit (provided it exists)

$$\lim_{k\to\infty} \sum_{(m,n)\in S_k} a_{mn} = \lim_{k\to\infty} \sum_{m=1}^{k}\sum_{n=1}^{k} a_{mn}$$

as "summing by squares", since as we already noted, $\sum_{(m,n)\in S_k} a_{mn}$ involves summing the a_{mn}'s within a $k \times k$ square.

Example 6.30. Now let

$$S_k = T_1 \cup \cdots \cup T_k \quad , \quad \text{where } T_\ell = \{(m,n)\,;\, m+n = \ell+1\}.$$

Notice that $T_\ell = \{(m,n)\,;\, m+n = \ell+1\} = \{(1,\ell),(2,\ell-1),\ldots,(\ell,1)\}$ represents the ℓ-th diagonal in the right-hand picture in Figure 6.5; for instance, $T_3 = \{(1,3),(2,2),(3,1)\}$ is the third diagonal in Figure 6.5. Then

$$\sum_{(m,n)\in S_k} a_{mn} = \sum_{\ell=1}^{k} \sum_{(m,n)\in T_\ell} a_{mn}$$

is the sum of the a_{mn}'s that are within the triangle consisting of the first k diagonals. It is natural to refer to the limit (provided it exists)

$$\lim_{k\to\infty} \sum_{(m,n)\in S_k} a_{mn} = \lim_{k\to\infty} \sum_{\ell=1}^{k} \sum_{(m,n)\in T_\ell} a_{mn},$$

as "summing by triangles". Using that $T_\ell = \{(1,\ell), (2,\ell-1), \ldots, (\ell,1)\}$, we can express the summation by triangles as

$$\sum_{k=1}^{\infty} \big(a_{1,k} + a_{2,k-1} + \cdots + a_{k,1}\big).$$

More generally, we can "sum by curves" as long as the curves increasingly fill up the array like the squares or triangles shown in Figure 6.5. More precisely, suppose that $S_1 \subseteq S_2 \subseteq S_3 \subseteq \cdots \subseteq \mathbb{N}\times\mathbb{N}$ is a nondecreasing sequence of finite sets having the property that for any m, n there is a k such that

$$\{1,2,\ldots,m\} \times \{1,2,\ldots,n\} \subseteq S_k \subseteq S_{k+1} \subseteq S_{k+2} \subseteq \cdots . \tag{6.23}$$

In the following theorem we consider the sequence $\{s_k\}$ where for each $k \in \mathbb{N}$, s_k is the finite sum

$$s_k := \sum_{(m,n)\in S_k} a_{mn}, \tag{6.24}$$

obtained by summing over all a_{mn} with (m,n) inside S_k.

Theorem 6.25 (**Sum by curves theorem**). *If a double series $\sum a_{mn}$ of complex numbers is absolutely convergent, then $\sum a_{mn}$ itself converges; moreover, the sequence $\{s_k\}$ defined in* (6.24) *converges, and*

$$\sum a_{mn} = \lim s_k.$$

Proof. We first show that $\{s_k\}$ is Cauchy and therefore converges, then we prove that $\sum a_{mn}$ converges and $\sum a_{mn} = \lim s_k$.

Step 1: To prove that $\{s_k\}$ is Cauchy, let $\varepsilon > 0$ be given. By assumption, $\sum |a_{mn}|$ converges, so if L denotes its limit and t_{mn} its m,n-th partial sum, we can choose N such that

$$m, n > N \quad \Longrightarrow \quad |L - t_{mn}| < \frac{\varepsilon}{2}. \tag{6.25}$$

Fix $n > N$. Then by the property (6.23) there is an $N' \in \mathbb{N}$ such that

$$\{1,2,\ldots,n\} \times \{1,2,\ldots,n\} \subseteq S_{N'} \subseteq S_{N'+1} \subseteq S_{N'+2} \subseteq \cdots . \tag{6.26}$$

Fix $k > \ell > N'$. Then, since $S_\ell \subseteq S_k$, we have

$$|s_k - s_\ell| = \Bigg| \sum_{(i,j)\in S_k} a_{ij} - \sum_{(i,j)\in S_\ell} a_{ij} \Bigg| = \Bigg| \sum_{(i,j)\in S_k\setminus S_\ell} a_{ij} \Bigg| \le \sum_{(i,j)\in S_k\setminus S_\ell} |a_{ij}|.$$

Since $\ell > N'$, by (6.26), $\{1,2,\ldots,n\} \times \{1,2,\ldots,n\} \subseteq S_\ell$. Now choose $m > n$ such that $S_k \subseteq \{1,2,\ldots,m\} \times \{1,2,\ldots,m\}$. Then,

$$\begin{aligned}\sum_{(i,j)\in S_k \setminus S_\ell} |a_{ij}| &\le \sum_{i=1}^{m}\sum_{j=1}^{m} |a_{ij}| \ - \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| \\ &= t_{mm} - t_{nn} \\ &= (t_{mm} - L) + (L - t_{nn}) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,\end{aligned}$$

where we used (6.25). Hence, $|s_k - s_\ell| < \varepsilon$ and so $\{s_k\}$ is Cauchy.

Step 2: We now show that $\sum a_{mn}$ converges with sum equal to $s := \lim s_k$. Let $\varepsilon > 0$ be given and choose N such that (6.25) holds with $\varepsilon/2$ replaced with $\varepsilon/3$. Fix natural numbers $m, n > N$. By the property (6.23) and the fact that $s_k \to s$ we can choose a $k > N$ such that

$$\{1,2,\ldots,m\} \times \{1,2,\ldots,n\} \subseteq S_k$$

and $|s_k - s| < \varepsilon/3$. Observe that

$$\begin{aligned}|s_k - s_{mn}| &= \left| \sum_{(i,j)\in S_k} a_{ij} \ - \sum_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}} a_{ij} \right| \\ &= \left| \sum_{(i,j)\in S_k\setminus\{1,\ldots,m\}\times\{1,\ldots,n\}} a_{ij} \right| \le \sum_{(i,j)\in S_k\setminus(\{1,\ldots,m\}\times\{1,\ldots,n\})} |a_{ij}|.\end{aligned}$$

Now choose $m' \in \mathbb{N}$ such that $S_k \subseteq \{1,2,\ldots,m'\} \times \{1,2,\ldots,m'\}$. Then,

$$\begin{aligned}\sum_{(i,j)\in S_k\setminus(\{1,\ldots,m\}\times\{1,\ldots,n\})} |a_{ij}| \ &\le \sum_{i=1}^{m'}\sum_{j=1}^{m'} |a_{ij}| \ - \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}| \\ &= t_{m'm'} - t_{mn} \\ &= (t_{m'm'} - L) + (L - t_{mn}) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \frac{2\varepsilon}{3},\end{aligned}$$

where we used the property (6.25) (with $\varepsilon/2$ replaced with $\varepsilon/3$). Finally, recalling that $|s_k - s| < \varepsilon/3$, by the triangle inequality, we have

$$|s_{mn} - s| \le |s_{mn} - s_k| + |s_k - s| < \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

This proves that $\sum a_{mn} = s$ and completes our proof. □

We recommend the reader to look at Exercise 11 for a related result.

6.5.3. Cauchy's double series theorem. Instead of summing by curves, in many applications we are interested in summing by rows or by columns.

THEOREM 6.26 (**Cauchy's double series theorem**). *For any double series $\sum a_{mn}$ of complex numbers, the following are equivalent statements:*

(a) The series $\sum a_{mn}$ is absolutely convergent;
(b) $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |a_{mn}|$ converges;
(c) $\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |a_{mn}|$ converges.

In either of these cases,

$$\sum a_{mn} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{mn} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{mn} \tag{6.27}$$

in the sense that both iterated sums converge and are equal to the sum of the series.

PROOF. We proceed in three steps.

Step 1: Assume first that the sum $\sum a_{mn}$ converges absolutely; we shall prove that both iterated sums $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}|a_{mn}|$, $\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}|a_{mn}|$ converge. Since $\sum|a_{mn}|$ converges, setting $s := \sum|a_{mn}|$ and denoting by s_{mn} the m,n-th partial sum, by definition of convergence we can choose N such that

$$m,n > N \implies |s - s_{mn}| < 1 \implies s_{mn} < s + 1. \tag{6.28}$$

Given $p \in \mathbb{N}$, choose $m \geq p$ such that $m > N$ and let $n > N$. Then in view of (6.28) we have

$$\sum_{j=1}^{n}|a_{pj}| \leq \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}| = s_{mn} < s + 1. \tag{6.29}$$

Therefore, the partial sums of $\sum_{j=1}^{\infty}|a_{pj}|$ are bounded above by a fixed constant and hence (by the nonnegative series test — see Theorem 3.20), for any $p \in \mathbb{N}$, the sum $\sum_{j=1}^{\infty}|a_{pj}|$ exists. Similarly, for each $q \in \mathbb{N}$, the sum $\sum_{i=1}^{\infty}|a_{iq}|$ exists. Therefore, by Pringsheim's theorem for series, both iterated series $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}|a_{mn}|$, $\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}|a_{mn}|$ converge (and equal $\sum|a_{mn}|$).

Step 2: Assuming that $\sum a_{mn}$ converges absolutely, we now establish the equality (6.27). Indeed, by the sum by curves theorem we know that $\sum a_{mn}$ converges and we showed in **Step 1** that for each $p,q \in \mathbb{N}$, the sums $\sum_{n=1}^{\infty}|a_{pn}|$ and $\sum_{m=1}^{\infty}|a_{mq}|$ exist. This implies that for each $p,q \in \mathbb{N}$, $\sum_{n=1}^{\infty}a_{pn}$ and $\sum_{m=1}^{\infty}a_{mq}$ converge. Now (6.27) follows from Pringsheim's theorem.

Step 3: Now assume that

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}|a_{mn}| = t < \infty.$$

We will show that $\sum a_{mn}$ is absolutely convergent; a similar proof shows that if $\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}|a_{mn}| < \infty$, then $\sum a_{mn}$ is absolutely convergent. Let $\varepsilon > 0$. Then the fact that $\sum_{i=1}^{\infty}\big(\sum_{j=1}^{\infty}|a_{ij}|\big) < \infty$ implies, by the Cauchy criterion for series, there is an N such that

$$k > m > N \implies \sum_{i=m+1}^{k}\Big(\sum_{j=1}^{\infty}|a_{ij}|\Big) < \frac{\varepsilon}{2}.$$

Let $m,n > N$. Then for any $k > m$, we have

$$\left|\sum_{i=1}^{k}\sum_{j=1}^{\infty}|a_{ij}| - \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\right| \leq \sum_{i=m+1}^{k}\sum_{j=1}^{\infty}|a_{ij}| < \frac{\varepsilon}{2}.$$

Taking $k \to \infty$ shows that for all $m,n > N$,

$$\left|t - \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\right| \leq \frac{\varepsilon}{2} < \varepsilon,$$

which proves that $\sum |a_{mn}|$ converges, and completes the proof of our result. □

Now for some double series examples.

Example 6.31. For our first example, consider the sum $\sum 1/(m^p n^q)$ where $p, q \in \mathbb{R}$. Since in this case,

$$\sum_{n=1}^{\infty} \frac{1}{m^p n^q} = \frac{1}{m^p} \cdot \Big(\sum_{n=1}^{\infty} \frac{1}{n^q} \Big),$$

it follows that

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{m^p n^q} = \Big(\sum_{m=1}^{\infty} \frac{1}{m^p} \Big) \cdot \Big(\sum_{n=1}^{\infty} \frac{1}{n^q} \Big).$$

Therefore, by Cauchy's double series theorem and the p-test, $\sum 1/(m^p n^q)$ converges if and only if both $p, q > 1$.

Example 6.32. The previous example can help us with other examples such as $\sum 1/(m^4 + n^4)$. Observe that

$$(m^2 - n^2)^2 \geq 0 \quad \Longrightarrow \quad m^4 + n^4 - 2m^2 n^2 \geq 0 \quad \Longrightarrow \quad \frac{1}{m^4 + n^4} \leq \frac{1}{2m^2 n^2}.$$

Since $\sum 1/(m^2 n^2)$ converges, by an easy generalization of our good ole comparison test (Theorem 3.27) to double series, we see that $\sum 1/(m^4 + n^4)$ converges too.

Example 6.33. For an application of Cauchy's theorem and the sum by curves theorem, we look at the double sum $\sum z^{m+n}$ for $|z| < 1$. For such z, this sum converges absolutely because

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} |z|^{m+n} = \sum_{m=0}^{\infty} |z|^m \cdot \frac{1}{1 - |z|} = \frac{1}{(1 - |z|)^2} < \infty,$$

where we used the geometric series test (twice): If $|r| < 1$, then $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$. So $\sum z^{m+n}$ converges absolutely by Cauchy's double series theorem, and

$$\sum z^{m+n} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} z^{m+n} = \sum_{m=0}^{\infty} z^m \cdot \frac{1}{1 - z} = \frac{1}{(1 - z)^2}.$$

On the other hand, by our sum by curves theorem, we can determine $\sum z^{m+n}$ by summing over curves; we shall choose to sum over triangles. Thus, if we set

$$S_k = T_0 \cup T_1 \cup T_2 \cup \cdots \cup T_k \quad , \quad \text{where } T_\ell = \{(m, n)\,;\, m + n = \ell \ , \ m, n \geq 0\},$$

then

$$\sum z^{m+n} = \lim_{k \to \infty} \sum_{(m,n) \in S_k} z^{m+n} = \lim_{k \to \infty} \sum_{\ell=0}^{k} \sum_{(m,n) \in T_\ell} z^{m+n}.$$

Since $T_\ell = \{(m, n)\,;\, m + n = \ell\} = \{(0, \ell), (1, \ell - 1), \ldots, (\ell, 0)\}$, we have

$$\sum_{(m,n) \in T_\ell} z^{m+n} = z^{0+\ell} + z^{1+(\ell-1)} + z^{2+(\ell-2)} + \cdots + z^{\ell+0} = (\ell + 1) z^\ell.$$

Thus, $\sum z^{m+n} = \sum_{k=0}^{\infty} (k + 1) z^k$. However, we already proved that $\sum z^{m+n} = 1/(1 - z)^2$, so

$$\frac{1}{(1 - z)^2} = \sum_{n=1}^{\infty} n z^{n-1}. \tag{6.30}$$

See Problem 4 for an easier proof of (6.30) using Cauchy's double series theorem.

Example 6.34. Another very neat application of Cauchy's double series theorem is to derive nonobvious identities. For example, let $|z| < 1$ and consider the series

$$\sum_{n=1}^{\infty} \frac{z^n}{1+z^{2n}} = \frac{z}{1+z^2} + \frac{z^2}{1+z^4} + \frac{z^3}{1+z^6} + \cdots ;$$

we'll see why this converges in a moment. Observe that (since $|z| < 1$)

$$\frac{1}{1+z^{2n}} = \sum_{m=0}^{\infty} (-1)^m z^{2mn},$$

by the familiar geometric series test with $r = -z^{2n}$: Since $|r| < 1$, then $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$. Therefore,

$$\sum_{n=1}^{\infty} \frac{z^n}{1+z^{2n}} = \sum_{n=1}^{\infty} z^n \cdot \sum_{m=0}^{\infty} (-1)^m z^{2mn} = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} (-1)^m z^{(2m+1)n}$$

We claim that the double sum $\sum (-1)^m z^{(2m+1)n}$ converges absolutely. To prove this, observe that

$$\sum_{n=1}^{\infty} \sum_{m=0}^{\infty} |z|^{(2m+1)n} = \sum_{n=1}^{\infty} |z|^n \sum_{m=0}^{\infty} |z|^{2nm} = \sum_{n=1}^{\infty} \frac{|z|^n}{1-|z|^{2n}}.$$

Since $\frac{1}{1-|z|^{2n}} \le \frac{1}{1-|z|}$ (this is because $|z|^{2n} \le |z|$ for $|z| < 1$), we have

$$\frac{|z|^n}{1-|z|^{2n}} \le \frac{1}{1-|z|} \cdot |z|^n.$$

Since $\sum |z|^n$ converges, by the comparison theorem, $\sum_{n=1}^{\infty} \frac{|z|^n}{1-|z|^{2n}}$ converges too. Hence, Cauchy's double series theorem applies, and

$$\begin{aligned}
\sum_{n=1}^{\infty} \sum_{m=0}^{\infty} (-1)^m z^{(2m+1)n} &= \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} (-1)^m z^{(2m+1)n} \\
&= \sum_{m=0}^{\infty} (-1)^m \sum_{n=1}^{\infty} z^{(2m+1)n} \\
&= \sum_{m=0}^{\infty} (-1)^m \frac{z^{2m+1}}{1-z^{2m+1}}.
\end{aligned}$$

Thus,

$$\sum_{n=1}^{\infty} \frac{z^n}{1+z^{2n}} = \sum_{m=0}^{\infty} (-1)^m \frac{z^{2m+1}}{1-z^{2m+1}};$$

that is, we have derived the striking identity between even and odd powers of z:

$$\frac{z}{1+z^2} + \frac{z^2}{1+z^4} + \frac{z^3}{1+z^6} + \cdots = \frac{z}{1-z} - \frac{z^3}{1-z^3} + \frac{z^5}{1-z^5} - + \cdots .$$

There are more beautiful series like this found in the exercises (see Problem 5 or better yet, Problem 7). We just touch on one more because it's so nice:

6.5.4. A neat ζ-function identity. Recall that the ζ-function is defined by $\zeta(z) = \sum_{n=1}^\infty \frac{1}{n^z}$, which converges absolutely for $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$. Here's a beautiful theorem from Flajolet and Vardi [**75, 232**].

THEOREM 6.27. *If $f(z) = \sum_{n=2}^\infty a_n\, z^n$ and $\sum_{n=2}^\infty |a_n|$ converges, then*

$$\boxed{\sum_{n=1}^\infty f\Big(\frac{1}{n}\Big) = \sum_{n=2}^\infty a_n\, \zeta(n).}$$

PROOF. We first write

$$\sum_{n=1}^\infty f\Big(\frac{1}{n}\Big) = \sum_{n=1}^\infty \sum_{m=2}^\infty a_m\, \frac{1}{n^m}.$$

Now if we set $C := \sum_{m=2}^\infty |a_m| < \infty$, then

$$\sum_{n=1}^\infty \sum_{m=2}^\infty \Big| a_m\, \frac{1}{n^m} \Big| \le \sum_{n=1}^\infty \sum_{m=2}^\infty |a_m|\, \frac{1}{n^2} \le C \sum_{n=1}^\infty \frac{1}{n^2} < \infty.$$

Hence, by Cauchy's double series theorem, we can switch the order of summation:

$$\sum_{n=1}^\infty f\Big(\frac{1}{n}\Big) = \sum_{n=1}^\infty \sum_{m=2}^\infty a_m \frac{1}{n^m} = \sum_{m=2}^\infty a_m \sum_{n=1}^\infty \frac{1}{n^m} = \sum_{m=2}^\infty a_m\, \zeta(m),$$

which completes our proof. □

Using this theorem we can derive the pretty formula (see Problem 9):

$$\boxed{\log 2 = \sum_{n=2}^\infty \frac{1}{2^n}\, \zeta(n).} \tag{6.31}$$

Not only is this formula pretty, it converges to $\log 2$ much faster than the usual series $\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n}$ (from which (6.31) is derived by the help of Theorem 6.27); see [**75, 232**] for a discussion of such convergence issues.

EXERCISES 6.5.

1. Determine the convergence of the limits and the iterated limits for the double sequences

$$(a)\ s_{mn} = \frac{1}{m} + \frac{1}{n} \quad , \quad (b)\ s_{mn} = \frac{m}{m+n} \quad , \quad (c)\ s_{mn} = \Big(\frac{n+1}{n+2}\Big)^m,$$

$$(d)\ s_{mn} = (-1)^{m+n}\Big(\frac{1}{m} + \frac{1}{n}\Big) \quad , \quad (e)\ s_{mn} = \frac{1}{1+(m-n)^2}.$$

2. Determine the convergence, iterated convergence, and absolute convergence, for the double series

$$(a)\ \sum_{m,n\ge 1} \frac{(-1)^{mn}}{mn} \quad , \quad (b)\ \sum_{m,n\ge 1} \frac{(-1)^n}{(m+n^p)(m+n^p-1)}\ ,\ p>1 \quad , \quad (c)\ \sum_{m\ge 2, n\ge 1} \frac{1}{m^n}$$

   Suggestion: For (b), show that $\sum_{m=1}^\infty \frac{1}{(m+n^p)(m+n^p-1)}$ telescopes.

3. (**mn-term test for double series**) Show that if $\sum a_{mn}$ converges, then $a_{mn} \to 0$. Suggestion: First verify that $a_{mn} = s_{mn} - s_{m-1,n} - s_{m,n-1} + s_{m-1,n-1}$.
4. Let $z \in \mathbb{C}$ with $|z| < 1$. For $(m,n) \in \mathbb{N} \times \mathbb{N}$, define $a_{mn} = z^n$ if $m \le n$ and define $a_{mn} = 0$ otherwise. Using Cauchy's double series theorem on $\sum a_{mn}$, prove (6.30). Using (6.30), find $\sum_{n=1}^\infty \frac{n}{2^n}$ (cf. Problem 3 in Exercises 3.5).

5. Let $|z| < 1$. Using Cauchy's double series theorem, derive the beautiful identities

$$(a)\ \frac{z}{1+z^2} + \frac{z^3}{1+z^6} + \frac{z^5}{1+z^{10}} + \cdots = \frac{z}{1-z^2} - \frac{z^3}{1-z^6} + \frac{z^5}{1-z^{10}} - + \cdots,$$

$$(b)\ \frac{z}{1+z^2} - \frac{z^2}{1+z^4} + \frac{z^3}{1+z^6} - + \cdots = \frac{z}{1+z} - \frac{z^3}{1+z^3} + \frac{z^5}{1+z^5} - + \cdots,$$

$$(c)\ \frac{z}{1+z} - \frac{2z^2}{1+z^2} + \frac{3z^3}{1+z^3} - + \cdots = \frac{z}{(1+z)^2} - \frac{z^2}{(1+z^2)^2} + \frac{z^3}{(1+z^3)^2} - + \cdots.$$

Suggestion: For (c), you need the formula $1/(1-z)^2 = \sum_{n=1}^\infty nz^{n-1}$ found in (6.30).

6. Here's a neat formula for $\zeta(k)$ found in [**40**]: For any $k \in \mathbb{N}$ with $k \geq 3$, we have

$$\boxed{\zeta(k) = \sum_{\ell=1}^{k-2} \sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{m^\ell (m+n)^{k-\ell}}.}$$

To prove this you may proceed as follows.

(i) Show that

$$\sum_{\ell=1}^{k-2} \frac{1}{m^\ell (m+n)^{k-\ell}} = \frac{1}{(m+n)^k} \sum_{\ell=1}^{k-2} \left(\frac{m+n}{m}\right)^\ell = \frac{1}{m^{k-2} n(m+n)} - \frac{1}{n(m+n)^{k-1}}.$$

(ii) Use (i) to show that

$$\sum_{\ell=1}^{k-2} \sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{m^\ell (m+n)^{k-\ell}} = \sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{m^{k-2} n(m+n)} - \sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{n(m+n)^{k-1}}.$$

Make sure you justify each step; in particular, why does each sum converge?

(iii) Use the partial fractions $\frac{1}{n(m+n)} = \frac{1}{n} - \frac{1}{m+n}$ to show that

$$\sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{m^{k-2} n(m+n)} = \sum_{m=1}^\infty \frac{1}{m^{k-1}} \sum_{n=1}^m \frac{1}{n}.$$

(iv) Replace the summation variable n with $\ell = m + n$ in $\sum_{m=1}^\infty \sum_{n=1}^\infty \frac{1}{n(m+n)^{k-1}}$ to get a new sum in terms of m and ℓ, then use Cauchy's double series theorem to change the order of summation. Finally, prove the desired result.

7. (**Number theory series**) Here are some pretty formulas involving number theory!

(a) For $n \in \mathbb{N}$, let $\tau(n)$ denote the number of positive divisors of n (that is, the number of positive integers that divide n). For example, $\tau(1) = 1$ and $\tau(4) = 3$ (because $1, 2, 4$ divide 4). Prove that

$$\sum_{n=1}^\infty \frac{z^n}{1-z^n} = \sum_{n=1}^\infty \tau(n) z^n \quad , \quad |z| < 1. \tag{6.32}$$

Suggestion: Write $1/(1-z^n) = \sum_{m=0}^\infty z^{mn} = \sum_{m=1}^\infty z^{n(m-1)}$, then prove that the left-hand side of (6.32) equals $\sum z^{mn}$. Finally, use Theorem 6.25 with the set S_k given by $S_k = T_1 \cup \cdots \cup T_k$ where $T_k = \{(m,n) \in \mathbb{N} \times \mathbb{N}\,;\, m \cdot n = k\}$.

(b) For $n \in \mathbb{N}$, let $\sigma(n)$ denote the sum of the positive divisors of n. For example, $\sigma(1) = 1$ and $\sigma(4) = 1 + 2 + 4 = 7$). Prove that

$$\sum_{n=1}^\infty \frac{z^n}{(1-z^n)^2} = \sum_{n=1}^\infty \sigma(n) z^n \quad , \quad |z| < 1.$$

8. Here is a neat problem. Let $f(z) = \sum_{n=1}^\infty a_n z^n$ and $g(z) = \sum_{n=1}^\infty b_n z^n$. Determine a set of points $z \in \mathbb{C}$ for which the following formula is valid:

$$\sum_{n=1}^\infty b_n f(z^n) = \sum_{n=1}^\infty a_n g(z^n).$$

From this formula, derive the following pretty formulas:

$$\sum_{n=1}^{\infty} f(z^n) = \sum_{n=1}^{\infty} \frac{a_n z^n}{1 - z^n} \quad , \quad \sum_{n=1}^{\infty} (-1)^{n-1} f(z^n) = \sum_{n=1}^{\infty} \frac{a_n z^n}{1 + z^n},$$

and my favorite:

$$\boxed{\sum_{n=1}^{\infty} \frac{f(z^n)}{n!} = \sum_{n=1}^{\infty} a_n e^{z^n}.}$$

9. In this problem we derive (6.31).
   (i) Prove that $\log 2 = \sum_{n=1}^{\infty} \frac{1}{2n(2n-1)} = \sum_{n=1}^{\infty} f\left(\frac{1}{n}\right)$, where $f(z) = \frac{z^2}{2(2-z)}$.
   (ii) Show that $f(z) = \sum_{n=2}^{\infty} \frac{z^n}{2^n}$ and from this and Theorem 6.27 prove (6.31).
10. (Cf. [**75, 232**]) Prove the following extension of Theorem 6.27: If $f(z) = \sum_{n=2}^{\infty} a_n\, z^n$ and for some $N \in \mathbb{N}$, $\sum_{n=2}^{\infty} \frac{|a_n|}{N^n}$ converges, then

$$\boxed{\sum_{n=N}^{\infty} f\left(\frac{1}{n}\right) = \sum_{n=2}^{\infty} a_n \left\{\zeta(n) - \left(1 + \frac{1}{2^n} + \cdots + \frac{1}{(N-1)^n}\right)\right\},}$$

where the sum $\left(1 + \frac{1}{2^n} + \cdots + \frac{1}{(N-1)^n}\right)$ is (by convention) zero if $N = 1$.
11. (**Arbitrary rearrangements of double series**) Let $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ be a bijective function and denote by $\nu_n = f(n) \in \mathbb{N} \times \mathbb{N}$; therefore $\nu_1, \nu_2, \nu_3, \ldots$ is a list of all elements of $\mathbb{N} \times \mathbb{N}$. For a double series $\sum a_{mn}$ of complex numbers, prove that $\sum_{n=1}^{\infty} a_{\nu_n}$ is absolutely convergent if and only if $\sum a_{mn}$ is absolutely convergent, in which case, $\sum_{n=1}^{\infty} a_{\nu_n} = \sum a_{mn}$.

6.6. Rearrangements and multiplication of power series

We already know that the associative law holds for infinite series. That is, we can group the terms of an infinite series in any way we wish and the resulting series still converges with the same sum (see Theorem 3.23). A natural question that you may ask is whether or not the commutative law holds for infinite series. That is, suppose that $s = a_1 + a_2 + a_3 + \cdots$ exists. Can we commute the a_n's in any way we wish and still get the same sum? For instance, is it true that

$$s = a_1 + a_2 + a_4 + a_3 + a_6 + a_8 + a_5 + a_{10} + a_{12} + \cdots ?$$

For general series, the answer is, quite shocking at first, "no!"

6.6.1. Rearrangements.

A sequence $\nu_1, \nu_2, \nu_3, \ldots$ of natural numbers such that every natural number occurs exactly once in this list is called a **rearrangement** of the natural numbers.

Example 6.35. $1, 2, 4, 3, 6, 8, 5, 10, 12, \ldots$, where we follow every odd number by two adjacent even numbers, is a rearrangement.

A **rearrangement** of a series $\sum_{n=1}^{\infty} a_n$ is a series $\sum_{n=1}^{\infty} a_{\nu_n}$ where $\{\nu_n\}$ is a rearrangement of $\mathbb{N}$.

Example 6.36. Let us rearrange the alternating harmonic series

$$\log 2 = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + - \cdots$$

using the rearrangement $1, 2, 4, 3, 6, 8, 5, 10, 12, \ldots$ we've already mentioned:

$$s = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + - - \\ \cdots + \frac{1}{2k-1} - \frac{1}{4k-2} - \frac{1}{4k} + \cdots ,$$

provided of course that this sum converges. Here, the bottom three terms represent the general formula for the k-th triplet of a positive term followed by two negative ones. To see that this sum converges, let s_n denote its n-th partial sum. Then we can write $n = 3k + \ell$ where ℓ is either 0, 1, or 2, and so

$$s_n = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + - - \cdots + \frac{1}{2k-1} - \frac{1}{4k-2} - \frac{1}{4k} + r_n,$$

where r_n consists of the next ℓ $(= 0, 1, 2)$ terms of the series for s_n. Note that $r_n \to 0$ as $n \to \infty$. In any case, we can write

$$\begin{aligned} s_n &= \left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + - - \cdots + \left(\frac{1}{2k-1} - \frac{1}{4k-2}\right) - \frac{1}{4k} + r_n \\ &= \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + - \cdots + \frac{1}{4k-2} - \frac{1}{4k} + r_n \\ &= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots + \frac{1}{2k-1} - \frac{1}{2k}\right) + r_n. \end{aligned}$$

Taking $n \to \infty$, we see that

$$s = \frac{1}{2} \log 2.$$

Thus, the rearrangement s has a different sum than the original series!

In summary, rearrangements of series can, in general, have different sums that the original series. In fact, it turns out that a convergent series can be rearranged to get a different value if and only if the series is not absolutely convergent. The "only if" portion is proved in Theorem 6.29 and the "if" portion is proved in

6.6.2. Riemann's rearrangement theorem.

Theorem 6.28 (**Riemann's rearrangement theorem**). *If a series $\sum a_n$ of real numbers converges, but not absolutely, then there are rearrangements of the series that can be made to converge to $\pm\infty$ or any real number whatsoever.*

Proof. We shall prove that there are rearrangements of the series that converge to any real number whatsoever; following the argument for this case, you should be able to handle the $\pm\infty$ cases yourself.

Step 1: We first show that the series corresponding to the positive and negative terms in $\sum a_n$ each diverge. Let $b_1, b_2, b_3, \ldots$ denote the terms in the sequence $\{a_n\}$ that are nonnegative, in the order in which they occur, and let $c_1, c_2, c_3, \ldots$ denote the absolute values of the terms in $\{a_n\}$ that are negative, again, in the order in which they occur. We claim that both series $\sum b_n$ and $\sum c_n$ diverge. To see this, observe that

$$\sum_{k=1}^{n} a_k = \sum_i b_i - \sum_j c_j, \tag{6.33}$$

where the right-hand sums are only over those natural numbers i, j such that b_i and c_j occur in the left-hand sum. The left-hand side converges as $n \to \infty$ by assumption, so if either sum $\sum_{n=1}^\infty b_n$ or $\sum_{n=1}^\infty c_n$ of nonnegative numbers converges, then the equality (6.33) would imply that the other sum converges. But this would then imply that

$$\sum_{k=1}^{n} |a_k| = \sum_i b_i + \sum_j c_j$$

converges as $n \to \infty$, which does not. Hence, both sums $\sum b_n$ and $\sum c_n$ diverge.

Step 2: We produce a rearrangement. Let $\xi \in \mathbb{R}$. We shall produce a rearrangement

$$\begin{aligned}(6.34)\quad b_1 + \cdots + b_{m_1} - c_1 - \cdots - c_{n_1} + b_{m_1+1} + \cdots + b_{m_2}\\ - c_{n_1+1} - \cdots - c_{n_2} + b_{m_2+1} + \cdots + b_{m_3} - c_{n_2+1} - \cdots\end{aligned}$$

such that its partial sums converge to ξ. We do so as follows. Let $\{\beta_n\}$ and $\{\gamma_n\}$ denote the partial sums for $\sum b_n$ and $\sum c_n$, respectively. Since $\beta_n \to \infty$, for n sufficiently large, $\beta_n > \xi$. We define m_1 as the smallest natural number such that

$$\beta_{m_1} > \xi.$$

Note that β_{m_1} differs from ξ by at most b_{m_1}. Since $\gamma_n \to \infty$, for n sufficiently large, $\beta_{m_1} - \gamma_n < \xi$. We define n_1 to be the smallest natural number such that

$$\beta_{m_1} - \gamma_{n_1} < \xi.$$

Note that the left-hand side differs from ξ by at most c_{n_1}. Now define m_2 as the smallest natural number greater than m_1 such that

$$\beta_{m_2} - \gamma_{n_1} > \xi.$$

As before, such a number exists because $\beta_n \to \infty$, and the left-hand side differs from ξ by at most b_{m_2}. We define the number n_2 as the smallest natural number greater than n_1 such that

$$\beta_{m_2} - \gamma_{n_2} < \xi,$$

where the left-hand side differs from ξ by at most c_{n_2}. Continuing this process, we produce sequences $m_1 < m_2 < m_3 < \cdots$ and $n_1 < n_2 < n_3 < \cdots$ such that for every k,

$$\beta_{m_k} - \gamma_{n_{k-1}} > \xi,$$

where the left-hand side differs from ξ by at most b_{m_k}, and

$$\beta_{m_k} - \gamma_{n_k} < \xi,$$

where the left-hand side differs from ξ by at most c_{n_k}.

Step 3: We now show that the series (6.34), which is just a rearrangement of $\sum a_n$, converges to ξ. Let

$$\begin{aligned}\beta'_k := b_1 + \cdots + b_{m_1} - c_1 - \cdots - c_{n_1} + b_{m_1+1} + \cdots + b_{m_2} -\\ \cdots - c_{n_{k-2}+1} - \cdots - c_{n_{k-1}} + b_{m_{k-1}+1} + \cdots + b_{m_k} = \beta_{m_k} - \gamma_{n_{k-1}}\end{aligned}$$

and

$$\begin{aligned}\gamma'_k := b_1 + \cdots + b_{m_1} - c_1 - \cdots - c_{n_1} + b_{m_1+1} + \cdots + b_{m_2} -\\ \cdots + b_{m_{k-1}+1} + \cdots + b_{m_k} - c_{n_{k-1}+1} - \cdots - c_{n_k} = \beta_{m_k} - \gamma_{n_k}.\end{aligned}$$

Then any given partial sum t of (6.34) is one of the following two sorts:

$$t = b_1 + \cdots + b_{m_1} - c_1 - \cdots - c_{n_1} + b_{m_1+1} + \cdots + b_{m_2} - \\ \cdots - c_{n_{k-2}+1} - \cdots - c_{n_{k-1}} + b_{m_{k-1}+1} + \cdots + b_\ell,$$

where $\ell \leq m_k$, in which case, $\gamma'_{k-1} < t \leq \beta'_k$; otherwise,

$$t = b_1 + \cdots + b_{m_1} - c_1 - \cdots - c_{n_1} + b_{m_1+1} + \cdots + b_{m_2} - \\ \cdots + b_{m_{k-1}+1} + \cdots + b_{m_k} - c_{n_{k-1}+1} - \cdots - c_\ell,$$

where $\ell \leq n_k$, in which case, $\gamma'_k \leq t < \beta'_k$. Now by construction, β'_k differs from ξ by at most b_{m_k} and γ'_k differs from ξ by at most c_{n_k}. Therefore, the fact that $\gamma'_{k-1} < t \leq \beta'_k$ or $\gamma'_k \leq t < \beta'_k$ imply that

$$\xi - c_{n_{k-1}} < t < \xi + b_{n_k} \quad \text{or} \quad \xi - c_{n_k} < t < \xi + b_{n_k}.$$

By assumption, $\sum a_n$ converges, so $b_{n_k}, c_{n_k} \to 0$, hence the partial sums of (6.34) must converge to ξ. This completes our proof. □

We now prove that a convergent series can be rearranged to get a different value only if the series is not absolutely convergent. Actually, we shall prove the contrapositive: If a series is absolutely convergent, then any rearrangement has the same value as the original sum. This is a consequence of the following theorem.

THEOREM 6.29 (**Dirichlet's theorem**). *All rearrangements of an absolutely convergent series of complex numbers converge with the same sum as the original series.*

PROOF. Let $\sum a_n$ converge absolutely. We shall prove that any rearrangement of this series converges to the same value as the sum itself. To see this, let $\nu_1, \nu_2, \nu_3, \ldots$ be any rearrangement of the natural numbers and define

$$a_{mn} = \begin{cases} a_m & \text{if } m = \nu_n, \\ 0 & \text{else.} \end{cases}$$

Then by definition of a_{mn}, we have

$$a_m = \sum_{n=1}^\infty a_{mn} \quad \text{and} \quad a_{\nu_n} = \sum_{m=1}^\infty a_{mn}.$$

Moreover,

$$\sum_{m=1}^\infty \sum_{n=1}^\infty |a_{mn}| = \sum_{m=1}^\infty |a_m| < \infty,$$

so by Cauchy's double series theorem,

$$\sum_{m=1}^\infty a_m = \sum_{m=1}^\infty \sum_{n=1}^\infty a_{mn} = \sum_{n=1}^\infty \sum_{m=1}^\infty a_{mn} = \sum_{n=1}^\infty a_{\nu_n}.$$

□

We now move to the important topic of multiplication of series.

6.6.3. Multiplication of power series and infinite series. If we consider two power series $\sum_{n=0}^\infty a_n z^n$ and $\sum_{n=0}^\infty b_n z^n$, then *formally* multiplying and combining like powers of z, we get

$$\begin{aligned}\big(a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots\big)\big(b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \cdots\big) = \\ a_0 b_0 + (a_0 b_1 + a_1 b_0)z + (a_0 b_2 + a_1 b_1 + a_2 b_0)z^2 \\ + (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0)z^3 + \cdots .\end{aligned}$$

In particular, taking $z = 1$, we get (again, only formally!)

$$\begin{aligned}\big(a_0 + a_1 + a_2 + a_3 + \cdots\big)\big(b_0 + b_1 + b_2 + b_3 + \cdots\big) = \\ a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) \\ + (a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0) + \cdots .\end{aligned}$$

These thoughts suggest the following definition. Given two series $\sum_{n=0}^\infty a_n$ and $\sum_{n=0}^\infty b_n$, their **Cauchy product** is the series $\sum_{n=0}^\infty c_n$, where

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0 = \sum_{k=0}^n a_k b_{n-k}.$$

A natural question to ask is if $\sum_{n=0}^\infty a_n$ and $\sum_{n=0}^\infty b_n$ converge, then is it true that

$$\Big(\sum_{n=0}^\infty a_n\Big)\Big(\sum_{n=0}^\infty b_n\Big) = \sum_{n=0}^\infty c_n \ ?$$

The answer is, what may be a surprising, "no".

Example 6.37. Let us consider the example $(\sum_{n=1}^\infty \frac{(-1)^{n-1}}{\sqrt{n}})(\sum_{n=1}^\infty \frac{(-1)^{n-1}}{\sqrt{n}})$, which is due to Cauchy. That is, let $a_0 = b_0 = 0$ and

$$a_n = b_n = (-1)^{n-1}\frac{1}{\sqrt{n}}, \quad n = 1, 2, 3, \ldots .$$

We know, by the alternating series test, that $\sum_{n=1}^\infty \frac{(-1)^{n-1}}{\sqrt{n}}$ converges. However, we shall see that the Cauchy product does not converge. Indeed,

$$c_0 = a_0 b_0 = 0, \quad c_1 = a_0 b_1 + a_1 b_0 = 0,$$

and for $n \geq 2$,

$$c_n = \sum_{k=0}^n a_k b_{n-k} = \sum_{k=1}^{n-1} \frac{(-1)^k (-1)^{n-k}}{\sqrt{k}\sqrt{n-k}} = (-1)^n \sum_{k=1}^{n-1} \frac{1}{\sqrt{k}\sqrt{n-k}}.$$

Since for $1 \leq k \leq n-1$, we have

$$k(n-k) \leq (n-1)(n-1) = (n-1)^2 \quad \Longrightarrow \quad \frac{1}{n-1} \leq \frac{1}{\sqrt{k(n-k)}},$$

we see that

$$(-1)^n c_n = \sum_{k=1}^{n-1} \frac{1}{\sqrt{k(n-k)}} \geq \sum_{k=1}^{n-1} \frac{1}{n-1} = \frac{1}{n-1}\sum_{k=1}^{n-1} 1 = 1.$$

Thus, the terms c_n do not tend to zero as $n \to \infty$, so by the n-th term test, the series $\sum_{n=0}^\infty c_n$ does not converge.

The problem with this example is that the series $\sum \frac{(-1)^{n-1}}{\sqrt{n}}$ does not converge absolutely. However, for absolutely convergent series, there is no problem as the following theorem, due to Franz Mertens (1840–1927), shows.

Theorem 6.30 (**Mertens' multiplication theorem**). *If at least one of two convergent series $\sum a_n = A$ and $\sum b_n = B$ converges absolutely, then their Cauchy product converges with sum equal to AB*

Proof. Consider the partial sums of the Cauchy product:

$$\begin{aligned} C_n &= c_0 + c_1 + \cdots + c_n \\ &= a_0b_0 + (a_0b_1 + a_1b_0) + \cdots + (a_0b_n + a_1b_{n-1} + \cdots + a_nb_0) \\ &= a_0(b_0 + \cdots + b_n) + a_1(b_0 + \cdots + b_{n-1}) + \cdots + a_nb_0. \end{aligned} \tag{6.35}$$

We need to show that C_n tends to AB as $n \to \infty$. Because our notation is symmetric in A and B, we may assume that the sum $\sum a_n$ is absolutely convergent. If A_n denotes the n-th partial sum of $\sum a_n$ and B_n that of $\sum b_n$, then from (6.35), we have

$$C_n = a_0B_n + a_1B_{n-1} + \cdots + a_nB_0.$$

If we set $B_k = B + \beta_k$, then $\beta_k \to 0$, and we can write

$$\begin{aligned} C_n &= a_0(B + \beta_n) + a_1(B + \beta_{n-1}) + \cdots + a_n(B + \beta_0) \\ &= A_nB + (a_0\beta_n + a_1\beta_{n-1} + \cdots + a_n\beta_0). \end{aligned}$$

Since $A_n \to A$, the first part of this sum converges to AB. Thus, we just need to show that the term in parenthesis tends to zero as $n \to \infty$. To see this, let $\varepsilon > 0$ be given. Putting $\alpha = \sum |a_n|$ and using that $\beta_n \to 0$, we can choose a natural number N such that for all $n > N$, we have $|\beta_n| < \varepsilon/(2\alpha)$. Also, since $\beta_n \to 0$, we can choose a constant C such that $|\beta_n| \le C$ for every n. Then for $n > N$,

$$\begin{aligned} |a_0\beta_n + a_1\beta_{n-1} + \cdots + a_n\beta_0| &= |a_0\beta_n + a_1\beta_{n-1} + \cdots + a_{n-N+1}\beta_{N+1} \\ &\qquad + a_{n-N}\beta_N + \cdots + a_n\beta_0| \\ &\le |a_0\beta_n + a_1\beta_{n-1} + \cdots + a_{n-N+1}\beta_{N+1}| + |a_{n-N}\beta_N + \cdots + a_n\beta_0| \\ &< \Big(|a_0| + |a_1| + \cdots + |a_{n-N+1}|\Big) \cdot \frac{\varepsilon}{2\alpha} + \Big(|a_{n-N}| + \cdots + |a_n|\Big) \cdot C \\ &\le \alpha \cdot \frac{\varepsilon}{2\alpha} + C\Big(|a_{n-N}| + \cdots + |a_n|\Big) \\ &= \frac{\varepsilon}{2} + C\Big(|a_{n-N}| + \cdots + |a_n|\Big). \end{aligned}$$

Since $\sum |a_n| < \infty$, by the Cauchy criterion for series, we can choose $N' > N$ such that

$$n > N' \quad \Longrightarrow \quad |a_{n-N}| + \cdots + |a_n| < \frac{\varepsilon}{2C}.$$

Then for $n > N'$, we see that

$$|a_0\beta_n + a_1\beta_{n-1} + \cdots + a_n\beta_0| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof of the theorem. □

As an easy corollary, we see that if $\sum_{n=0}^\infty a_n z^n$ and $\sum_{n=0}^\infty b_n z^n$ have radii of convergence R_1, R_2, respectively, then since power series converge absolutely within their radii of convergence, for all $z \in \mathbb{C}$ with $|z| < R_1, R_2$, we have

$$\Big(\sum_{n=0}^\infty a_n z^n \Big) \Big(\sum_{n=0}^\infty b_n z^n \Big) = \sum_{n=0}^\infty c_n z^n$$

where $c_n = \sum_{k=0}^n a_k b_{n-k}$. In words: *The product of power series is a power series.*

Here's a question: Suppose that $\sum a_n$ and $\sum b_n$ converge and their Cauchy product $\sum c_n$ also converges; is it true that $\sum c_n = \big(\sum a_n\big)\big(\sum b_n\big)$? The answer may seem to be an "obvious" yes. However, it's not so "obvious' because the definition of the Cauchy product was based on a formal argument. Here is a proof of this "obvious" fact.

Theorem 6.31 (**Abel's multiplication theorem**). *If the Cauchy product of two convergent series $\sum a_n = A$ and $\sum b_n = B$ converges, then the Cauchy product has the value AB.*

Proof. In my opinion, the slickest proof of this theorem is Abel's original, proved in 1826 [**120**, p. 321] using his limit theorem, Theorem 6.20. Let

$$f(z) = \sum a_n z^n, \quad g(z) = \sum b_n z^n, \quad h(z) = \sum c_n z^n,$$

where $c_n = a_0 b_n + \cdots + a_n b_0$. These power series converge at $z = 1$, so they must have radii of convergence at least 1. In particular, each series converges absolutely for $|z| < 1$ and for these values of z according to according to Merten's theorem, we have

$$h(z) = f(z) \cdot g(z).$$

Since each of the sums $\sum a_n$, $\sum b_n$, and $\sum c_n$ converges, by Abel's limit theorem, the functions f, g, and h converge to A, B, and $C = \sum c_n$, respectively, as $z = x \to 1$ from the left. Thus,

$$C = \lim_{x \to 1-} h(x) = \lim_{x \to 1-} f(x) \cdot g(x) = A \cdot B.$$

□

Example 6.38. For example, let us square $\log 2 = \sum_{n=1}^\infty \frac{(-1)^{n-1}}{n}$. It turns out that it will be convenient to write $\log 2$ in two ways: $\log 2 = \sum_{n=1}^\infty a_n$ where $a_0 = 0$ and $a_n = \frac{(-1)^{n-1}}{n}$ for $n = 1, 2, \ldots$, and as $\log 2 = \sum_{n=0}^\infty b_n$ where $b_n = \frac{(-1)^n}{n+1}$. Thus, $c_0 = a_0 b_0 = 0$ and for $n = 1, 2, \ldots$, we see that

$$c_n = \sum_{k=0}^n a_k b_{n-k} = \sum_{k=1}^n \frac{(-1)^{k-1}(-1)^{n-k}}{k(n+1-k)} = (-1)^{n-1} \alpha_n,$$

where $\alpha_n = \sum_{k=1}^n \frac{1}{k(n+1-k)}$. By Abel's multiplication theorem, we have $(\log 2)^2 = \sum_{n=0}^\infty c_n = \sum_{n=1}^\infty (-1)^{n-1} \alpha_n$ as long as this latter sum converges. By the alternating series test, this sum converges if we can prove that $\{\alpha_n\}$ is nonincreasing and converges to zero. To prove these statements hold, observe that we can write

$$\frac{1}{k(n-k+1)} = \frac{1}{n+1}\Big(\frac{1}{k} + \frac{1}{n-k+1}\Big),$$

therefore

$$\begin{aligned}\alpha_n &= \frac{1}{1}\cdot\frac{1}{n}+\frac{1}{2}\cdot\frac{1}{n-1}+\frac{1}{3}\cdot\frac{1}{n-2}+\cdots+\frac{1}{n}\cdot\frac{1}{1}\\ &= \frac{1}{n+1}\Big[\Big(1+\frac{1}{n}\Big)+\Big(\frac{1}{2}+\frac{1}{n-1}\Big)+\Big(\frac{1}{3}+\frac{1}{n-2}\Big)+\cdots+\Big(\frac{1}{n}+\frac{1}{1}\Big)\Big].\end{aligned}$$

In the brackets there are two copies of $1+\frac{1}{2}+\cdots+\frac{1}{n}$. Thus,

$$\alpha_n = \frac{2}{n+1}H_n, \qquad \text{where } H_n := 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}.$$

It is common to use the notation H_n for the n-th partial sum of the harmonic series. Now, recall from Section 4.6.5 on the Euler-Mascheroni constant that $\gamma_n := H_n - \log n$ is bounded above by 1, so

$$\begin{aligned}\alpha_n = \frac{2}{n+1}(\gamma_n+\log n) \le \frac{2}{n+1}+2\frac{\log n}{n+1} &= \frac{2}{n+1}+2\cdot\frac{n}{n+1}\cdot\frac{1}{n}\log n\\ &= \frac{2}{n+1}+2\cdot\frac{n}{n+1}\cdot\log(n^{1/n}) \to 0+2\cdot 1\cdot\log 1 = 0\end{aligned}$$

as $n\to\infty$. Thus, $\alpha_n \to 0$. Moreover,

$$\begin{aligned}\alpha_n-\alpha_{n+1} = \frac{2}{n+1}H_n-\frac{2}{n+2}H_{n+1} &= \frac{2}{n+1}H_n-\frac{2}{n+2}\Big(H_n+\frac{1}{n+1}\Big)\\ &= \Big(\frac{2}{n+1}-\frac{2}{n+2}\Big)H_n-\frac{2}{(n+1)(n+2)}\\ &= \frac{2}{(n+1)(n+2)}H_n-\frac{2}{(n+1)(n+2)}\\ &= \frac{2}{(n+1)(n+2)}(H_n-1)\ge 0.\end{aligned}$$

Thus, $\alpha_n \ge \alpha_{n+1}$, so $\sum c_n = \sum(-1)^{n-1}\alpha_n$ converges. Hence, we have proved the following pretty formula:

$$\boxed{\begin{aligned}\frac{1}{2}\big(\log 2\big)^2 &= \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n+1}H_n\\ &= \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n+1}\Big(1+\frac{1}{2}+\cdots+\frac{1}{n}\Big)\ .\end{aligned}}$$

Our final theorem, Cauchy's multiplication theorem, basically says that we can multiply absolutely convergent series without worrying about anything. To introduce this theorem, note that if we have *finite* sums $\sum a_n$ and $\sum b_n$, then

$$\Big(\sum a_n\Big)\cdot\Big(\sum b_n\Big) = \sum a_m b_n,$$

where the sum on the right means to add over all such products $a_m b_n$ in any order we wish. One can ask if this holds true in the infinite series realm. The answer is "yes" if both series on the left are absolutely convergent.

THEOREM 6.32 (**Cauchy's multiplication theorem**). *If two series $\sum a_n = A$ and $\sum b_n = B$ converge absolutely, then the double series $\sum a_m b_n$ converges absolutely and has the value AB.*

Proof. Since

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}|a_m b_n| = \sum_{m=0}^{\infty}|a_m|\sum_{n=0}^{\infty}|b_n| = \Big(\sum_{m=0}^{\infty}|a_m|\Big)\Big(\sum_{n=0}^{\infty}|b_n|\Big) < \infty,$$

by Cauchy's double series theorem, the double series $\sum a_m b_n$ converges absolutely, and we can iterate the sums:

$$\sum a_m b_n = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_m b_n = \sum_{m=0}^{\infty} a_m \sum_{n=0}^{\infty} b_n = \Big(\sum_{m=0}^{\infty} a_m\Big)\Big(\sum_{n=0}^{\infty} b_n\Big) = A\cdot B.$$

□

We remark that Cauchy's multiplication theorem generalizes to a product of more than two absolutely convergent series.

6.6.4. The exponential function (again). Using Mertens' or Cauchy's multiplication theorem, we can give an alternative and quick proof of the formula $\exp(z)\exp(w) = \exp(z+w)$ for $z, w \in \mathbb{C}$, which was originally proved in Theorem 3.31 using a completely different method:

$$\begin{aligned}\exp(z)\exp(w) &= \Big(\sum_{n=0}^{\infty}\frac{z^n}{n!}\Big)\cdot\Big(\sum_{n=0}^{\infty}\frac{w^n}{n!}\Big)\\ &= \sum_{n=0}^{\infty}\Big(\sum_{k=0}^{n}\frac{z^k}{k!}\cdot\frac{w^{n-k}}{(n-k)!}\Big)\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}z^k w^{n-k}\Big)\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\sum_{k=0}^{n}\binom{n}{k}z^k w^{n-k}\Big) = \sum_{n=0}^{\infty}\frac{1}{n!}(z+w)^n = \exp(z+w),\end{aligned}$$

where we used the binomial theorem for $(z+w)^n$ in the last line.

Exercises 6.6.

1. Here are some alternating series problems:
   (a) Prove that

$$\frac{1}{1}+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\cdots+\frac{1}{4k-3}+\frac{1}{4k-1}-\frac{1}{2k}+\cdots=\frac{3}{2}\log 2.$$

   that is, we rearrange the alternating harmonic series so that two positive terms are followed by one negative one, otherwise keeping the ordering the same. Suggestion: Observe that

$$\begin{aligned}\frac{1}{2}\log 2 &= \frac{1}{2}-\frac{1}{4}+\frac{1}{6}-\frac{1}{8}+\frac{1}{9}-\frac{1}{10}+\cdots\\ &= 0+\frac{1}{2}+0-\frac{1}{4}+0+\frac{1}{6}+0-\frac{1}{8}+\cdots.\end{aligned}$$

   Add this term-by-term to the series for $\log 2$.

   (b) Prove that

$$\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}-\frac{1}{2}+\cdots+\frac{1}{8k-7}+\frac{1}{8k-5}+\frac{1}{8k-3}+\frac{1}{8k-1}-\frac{1}{2k}+\cdots=\frac{3}{2}\log 2;$$

   that is, we rearrange the alternating harmonic series so that four positive terms are followed by one negative one, otherwise keeping the ordering the same.

(c) What's wrong with the following argument?

$$\begin{aligned}1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\cdots &= \Big(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots\Big)\\ &\quad -2\Big(\frac{1}{2}+0+\frac{1}{4}+0+\frac{1}{6}+\cdots\Big)\\ &= \Big(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots\Big)-\Big(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots\Big)=0.\end{aligned}$$

2. Let $f(z)=\sum_{n=0}^\infty a_n\, z^n$ be absolutely convergent for $|z|<1$. Prove that for $|z|<1$, we have

$$\frac{f(z)}{1-z}=\sum_{n=0}^\infty (a_0+a_1+a_2+\cdots+a_n)z^n.$$

3. Using the previous problem, prove that for $z\in\mathbb{C}$ with $|z|<1$,

$$\frac{1}{(1-z)^2}=\sum_{n=0}^\infty (n+1)z^n;\quad \text{that is,}\ \Big(\sum_{n=0}^\infty z^n\Big)\cdot\Big(\sum_{n=0}^\infty z^n\Big)=\sum_{n=0}^\infty (n+1)z^n.$$

Using this formula, derive the neat looking formula: For $z\in\mathbb{C}$ with $|z|<1$,

$$\Big(\sum_{n=0}^\infty \cos n\theta\, z^n\Big)\cdot\Big(\sum_{n=0}^\infty \sin n\theta\, z^n\Big)=\frac{1}{2}\sum_{n=0}^\infty (n+1)\sin n\theta\, z^n. \tag{6.36}$$

Suggestion: Put $z=e^{i\theta}x$ with x real into the formula $(\sum_{n=0}^\infty z^n)\cdot(\sum_{n=0}^\infty z^n)=\sum_{n=0}^\infty (n+1)z^n$, then equate imaginary parts of both sides; this proves (6.36) for $z=x$ real and $|x|<1$. Why does (6.36) hold for $z\in\mathbb{C}$ with $|z|<1$?

4. Derive the beautiful formula: For $|z|<1$,

$$\Big(\sum_{n=1}^\infty \frac{\cos n\theta}{n}z^n\Big)\cdot\Big(\sum_{n=1}^\infty \frac{\sin n\theta}{n}z^n\Big)=\frac{1}{2}\sum_{n=2}^\infty \frac{H_n\sin n\theta}{n}z^n.$$

5. In this problem we prove the following fact: Let $f(z)=\sum_{n=0}^\infty a_n z^n$ be a power series with radius of convergence $R>0$ and let $\alpha\in\mathbb{C}$ with $|\alpha|<R$. Then we can write

$$f(z)=\sum_{n=0}^\infty b_n(z-\alpha)^n,$$

where this series converges absolutely for $|z-\alpha|<R-\alpha$.

(i) Show that

$$f(z)=\sum_{n=0}^\infty\sum_{m=0}^n a_n\binom{n}{m}\alpha^{n-m}(z-\alpha)^m. \tag{6.37}$$

(ii) Prove that

$$\sum_{n=0}^\infty\sum_{m=0}^n |a_n|\binom{n}{m}|\alpha|^{n-m}|z-\alpha|^m=\sum_{n=0}^\infty |a_n|\big(|z-\alpha|+|\alpha|\big)^m<\infty.$$

(iii) Verifying that you can change the order of summation in (6.37), prove the result.

6.7. ★ Proofs that $\sum 1/p$ diverges

We know that the harmonic series $\sum 1/n$ diverges. However, if we only sum over the squares, then we get the convergent sum $\sum 1/n^2$. Similarly, if we only sum over the cubes, we get the convergent sum $\sum 1/n^3$. One may ask: What if we sum only over all primes:

$$\sum\frac{1}{p}=\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\cdots,$$

do we get a convergent sum? We know that there are arbitrarily large gaps between primes (see Problem 1 in Exercises 2.4), so one may conjecture that $\sum 1/p$ converges. However, following [**23**], [**63**], [**164**] (cf. [**165**]), and [**130**] we shall prove that $\sum 1/p$ diverges! Other proofs can be found in the exercises. An expository article giving other proofs (cf. [**153**], [**51**]) on this fascinating divergent sum can be found in [**231**].

6.7.1. Proof I: Proof by multiplication and rearrangement. This is Bellman [**23**] and Dux's [**63**] argument. Suppose, for sake of contradiction, that $\sum 1/p$ converges. Then we can fix a prime number m such that $\sum_{p>m} 1/p < 1$. Let $2 < 3 < \cdots < m$ be the list of all prime numbers up to m. Given $N > m$, let P_N be the set of natural numbers greater than one and less than or equal to N all of whose prime factors are less than or equal to m, and let Q_N be the set of natural numbers greater than one and less than or equal to N all of whose prime factors are greater than m. Explicitly,

$$\begin{aligned} &k \in P_N \iff 1 < k \le N \text{ and } k = 2^i 3^j \cdots m^k, \quad \text{some } i, j, \ldots, k, \\ &\ell \in Q_N \iff 1 < \ell \le N \text{ and } \ell = p\,q \cdots r, \quad p, q, \ldots, r > m \text{ are prime.} \end{aligned} \tag{6.38}$$

In the product $p\,q \cdots r$, prime numbers may be repeated. Observe that any integer $1 < n \le N$ that is not in P_N or Q_N must have prime factors that are both less than or equal to m and greater than m, and hence can be factored in the form $n = k\,\ell$ where $k \in P_N$ and $\ell \in Q_N$. Thus, the finite sum

$$\sum_{k \in P_N} \frac{1}{k} + \sum_{\ell \in Q_N} \frac{1}{\ell} + \Big(\sum_{k \in P_N} \frac{1}{k}\Big)\Big(\sum_{\ell \in Q_N} \frac{1}{\ell}\Big) = \sum_{k \in P_N} \frac{1}{k} + \sum_{\ell \in Q_N} \frac{1}{\ell} + \sum_{k \in P_N, \ell \in Q_N} \frac{1}{k\,\ell},$$

contains every number of the form $1/n$ where $1 < n \le N$. (Of course, the resulting sum contains other numbers too.) In particular,

$$\sum_{k \in P_N} \frac{1}{k} + \sum_{\ell \in Q_N} \frac{1}{\ell} + \Big(\sum_{k \in P_N} \frac{1}{k}\Big)\Big(\sum_{\ell \in Q_N} \frac{1}{\ell}\Big) \ge \sum_{n=2}^{N} \frac{1}{n},$$

We shall prove that the finite sums on the left remain bounded as $N \to \infty$, which contradicts the fact that the harmonic series diverges.

To see that $\sum_{P_N} 1/k$ converges, note that each geometric series $\sum_{j=1}^{\infty} 1/p^j$ converges (absolutely since all the $1/p^j$ are positive) to a finite real number. Hence, by Cauchy's multiplication theorem (or rather its generalization to a product of more than two absolutely convergent series), we have

$$\left(\sum_{i=1}^{\infty} \frac{1}{2^i}\right)\left(\sum_{j=1}^{\infty} \frac{1}{3^j}\right) \cdots \left(\sum_{k=1}^{\infty} \frac{1}{m^k}\right) = \sum \frac{1}{2^i 3^j \cdots m^k}$$

is a finite real number, where the sum on the right is over all $i, j, \ldots, k = 1, 2, \ldots$. Using the definition of P_N in (6.38), we see that $\sum_{P_N} 1/k$ is bounded above by this finite real number uniformly in N. Thus, $\lim_{N\to\infty} \sum_{P_N} 1/k$ is finite.

We now prove that $\lim_{N\to\infty} \sum_{Q_N} 1/\ell$ is finite. To do so observe that since $\alpha := \sum_{p>m} 1/p < 1$ and all the $1/p$'s are positive, the sum $\sum_{p>m} 1/p$, in particular, converges absolutely. Hence, by Cauchy's multiplication theorem, we have

$$\alpha^2 = \left(\sum_{p>m} \frac{1}{p}\right)^2 = \sum_{p,q>m} \frac{1}{p\,q},$$

where the sum is over all primes $p, q > m$, and

$$\alpha^3 = \left(\sum_{p>m} \frac{1}{p}\right)^3 = \sum_{p,q,r>m} \frac{1}{p\,q\,r},$$

where the sum is over all primes $p, q, r > m$. We can continue this procedure showing that α^j is the sum $\sum 1/(p\,q \cdots r)$ where the sum is over all j-tuples of primes $p, q, \ldots, r$ all of which are strictly larger than m. By definition of Q_N in (6.38), it follows that the sum $\sum_{Q_N} 1/\ell$ is bounded by the number $\sum_{j=1}^{\infty} \alpha^j$, which is finite because $\alpha < 1$. Hence, the limit $\lim_{N\to\infty} \sum_{Q_N} 1/\ell$ is finite, and we have reached a contradiction.

6.7.2. An elementary number theory fact. Our next proof depends on the idea of square-free integers. A positive integer is said to be **square-free** if no squared prime divides it, that is, if a prime occurs in its prime factorization, then it occurs with multiplicity one. For instance, 1 is square-free because no squared prime divides it, $10 = 2 \cdot 5$ is square-free, but $24 = 2^3 \cdot 3 = 2^2 \cdot 2 \cdot 3$ is not square-free.

We claim that any positive integer can be written uniquely as the product of a square and a square-free integer. Indeed, let $n \in \mathbb{N}$ and let k be the largest natural number such that k^2 divides n. Then n/k^2 must be square-free, for if n/k^2 is divided by a squared prime p^2, then $pk > k$ divides n, which is not possible by definition of k. Thus, any positive integer n can be uniquely written as $n = k^2$ if n is a perfect square, or

$$n = k^2 \cdot p\,q \cdots r, \tag{6.39}$$

where $k \geq 1$ and where $p, q, \ldots, r$ are some primes less than or equal to n that occur with multiplicity one. Using the fact that any positive integer can be uniquely written as the product of a square and a square-free integer, we shall prove that $\sum 1/p$ diverges.

6.7.3. Proof II: Proof by comparison. Here is Niven's [**164, 165**] proof. We first prove that the product

$$\prod_{p<N} \left(1 + \frac{1}{p}\right)$$

diverges to ∞ as $N \to \infty$, where the product is over all primes less than N. Let $2 < 3 < \cdots < m$ be all the primes less than N. Consider the product

$$\prod_{p<N} \left(1 + \frac{1}{p}\right) = \left(1 + \frac{1}{2}\right)\left(1 + \frac{1}{3}\right) \cdots \left(1 + \frac{1}{m}\right).$$

For example, if $N = 5$, then

$$\prod_{p<5} \left(1 + \frac{1}{p}\right) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{2 \cdot 3}.$$

If $N = 6$, then

$$\begin{aligned}\prod_{p<6} \left(1 + \frac{1}{p}\right) &= \left(1 + \frac{1}{2}\right)\left(1 + \frac{1}{3}\right)\left(1 + \frac{1}{5}\right) \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{2 \cdot 3} + \frac{1}{2 \cdot 5} + \frac{1}{3 \cdot 5} + \frac{1}{2 \cdot 3 \cdot 5}.\end{aligned}$$

Using induction on N, we can always write

$$\prod_{p<N}\Big(1+\frac{1}{p}\Big) = 1 + \sum_{p<N}\frac{1}{p} + \sum_{p,q<N}\frac{1}{p\cdot q} + \cdots + \sum_{p,q,\ldots,r<N}\frac{1}{p\cdot q\cdots r},$$

where the k-th sum on the right is the sum over over all reciprocals of the form $\frac{1}{p_1\cdot p_2\cdots p_k}$ with $p_1,\ldots,p_k$ *distinct* primes less than N. Thus,

$$\prod_{p<N}\Big(1+\frac{1}{p}\Big)\cdot\sum_{k<N}\frac{1}{k^2} = \sum_{k<N}\frac{1}{k^2} + \sum_{k<N}\sum_{p<N}\frac{1}{k^2 p} + \sum_{k<N}\sum_{p,q<N}\frac{1}{k^2\cdot p\cdot q} + \cdots + \sum_{k<N}\sum_{p,q,\ldots,r<N}\frac{1}{k^2\cdot p\cdot q\cdots r}.$$

By our discussion on square-free numbers around (6.39), the right-hand side contains every number of the form $1/n$ where $n<N$ (and many other numbers too). In particular,

$$\prod_{p<N}\Big(1+\frac{1}{p}\Big)\cdot\sum_{k<N}\frac{1}{k^2} \geq \sum_{n<N}\frac{1}{n}. \tag{6.40}$$

From this inequality, we shall prove that $\sum 1/p$ diverges. To this end, we know that $\sum_{k=1}^\infty 1/k^2$ converges while $\sum_{n=1}^\infty 1/n$ diverges, so it follows that

$$\lim_{N\to\infty}\prod_{p<N}\Big(1+\frac{1}{p}\Big) = \infty.$$

To relate this product to the sum $\sum 1/p$, note that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \geq 1 + x$$

for $x\geq 0$ — in fact, this inequality holds for all $x\in\mathbb{R}$ by Theorem 4.29. Hence,

$$\prod_{p<N}\Big(1+\frac{1}{p}\Big) \leq \prod_{p<N}\exp(1/p) = \exp\Big(\sum_{p<N}\frac{1}{p}\Big).$$

Since the left-hand side increases without bound as $N\to\infty$, so must the sum $\sum_{p<N} 1/p$. This ends **Proof II**; see Problem 2 for a related proof.

6.7.4. Proof III: Another proof by comparison.

This is Gilfeather and Meister's argument [**130**]. The first step is to prove that for any natural number $N>1$, we have

$$\prod_{p<N}\frac{p}{p-1} \geq \sum_{n=1}^{N-1}\frac{1}{n}.$$

To prove this we shall prove that $\prod_{p<N}\Big(1-\frac{1}{p}\Big)^{-1}\to\infty$. To see this, observe that

$$\Big(1-\frac{1}{p}\Big)^{-1} = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots.$$

Let $2<3<\cdots<m$ be all the primes less than N. Then every natural number $n<N$ can be written in the form

$$n = 2^i\,3^j\,\cdots\,m^k$$

for some nonnegative integers $i, j, \ldots, k$. It follows that the product

$$\begin{aligned}\prod_{p<N}\Big(1-\frac{1}{p}\Big)^{-1} &= \Big(1-\frac{1}{2}\Big)^{-1}\Big(1-\frac{1}{3}\Big)^{-1}\cdots\Big(1-\frac{1}{m}\Big)^{-1}\\ &= \Big(1+\frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}\cdots\Big)\Big(1+\frac{1}{3}+\frac{1}{3^2}+\frac{1}{3^3}+\cdots\Big)\cdots\\ &\qquad\cdots\Big(1+\frac{1}{m}+\frac{1}{m^2}+\frac{1}{m^3}+\cdots\Big)\end{aligned}$$

after multiplying out using Cauchy's multiplication theorem (or rather its generalization to a product of more than two absolutely convergent series), contains all the numbers $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{N-1}$ (and of course, many more numbers too). Thus,

$$\prod_{p<N}\frac{p}{p-1} = \prod_{p<N}\Big(1-\frac{1}{p}\Big)^{-1} \geq \sum_{n=1}^{N-1}\frac{1}{n}, \tag{6.41}$$

which proves our first step. Now recall from (4.29) that for any natural number n, we have

$$\frac{1}{n+1} < \log(n+1) - \log n < \frac{1}{n}. \tag{6.42}$$

In particular, taking logarithms of both sides of (6.41), we get

$$\begin{aligned}\log\Big(\sum_{n=1}^{N-1}\frac{1}{n}\Big) &\leq \log\Big(\prod_{p<N}\frac{p}{p-1}\Big)\\ &= \sum_{p<N}\big(\log p - \log(p-1)\big) \leq \sum_{p<N}\frac{1}{p-1} \leq \sum_{p<N}\frac{2}{p},\end{aligned}$$

where we used that $p \leq 2(p-1)$ (this is because $n \leq 2(n-1)$ for all natural numbers $n > 1$). Since $\sum_{n=1}^{N-1} 1/n \to \infty$ as $N \to \infty$, $\log\big(\sum_{n=1}^{N-1} 1/n\big) \to \infty$ as $N \to \infty$ as well, so the sum $\sum 1/p$ must diverge.

Exercises 6.7.

1. Let $s_n = 1/2 + 1/3 + \cdots + 1/p_n$ (where p_n is the n-th prime) be the n-th partial sum of $\sum 1/p$. We know that $s_n \to \infty$ as $n \to \infty$. However, it turns out that $s_n \to \infty$ avoiding all integers! Prove this. Suggestion: Multiply s_n by $2 \cdot 3 \cdots p_{n-1}$.
2. Niven's proof can be slightly modified to avoid using the square-free fact. Derive the inequality (6.40) (which, as shown in the main text, implies that $\sum 1/p$ diverges) by proving that for any prime p,
$$\Big(1+\frac{1}{p}\Big)\cdot\sum_{k=0}^{n}\frac{1}{p^{2k}} = \sum_{k=0}^{2n+1}\frac{1}{p^k}.$$
3. Here is another proof that is similar to Gilfeather and Meister's argument where we replace the inequality (6.42) with the following argument.
   (i) Prove that
$$\frac{1}{1-x/2} \leq e^x \quad \text{for all } 0 \leq x \leq 1. \tag{6.43}$$
   Suggestion: Prove that $e^{-x} \leq 1 - x/2$ using the series expansion for e^{-x}.
   (ii) Taking logarithms of (6.43), prove that for any prime number p, we have
$$-\log\Big(1-\frac{1}{p}\Big) = -\log\Big(1-\frac{2/p}{2}\Big) \leq \frac{2}{p}.$$

(iii) Prove that

$$\frac{1}{2}\sum_{p<N}\log\Big(\frac{p}{p-1}\Big) \le \sum_{p<N}\frac{1}{p}.$$

(iv) Finally, use (6.41) as in the main text to prove that $\sum 1/p$ diverges.

4. Here's Vanden Eynden's proof [**231**]. Assume that $\sum 1/p$ converges. Then we can choose an N such that $\alpha := \sum_{p>N} 1/p < 1/2$.

(i) For $x \ge 1$, let M_x be the set of all natural numbers $1 \le n \le x$ such that $n = 1$ or $n = p_1 \cdots p_k$ where the p_j's are prime and $p_j > N$. Prove that

$$\lim_{x\to\infty}\sum_{n\in M_x}\frac{1}{n} = \infty.$$

(ii) By (i), we can choose x such that $\beta := \sum_{n\in M_x}\frac{1}{n} > 2$. Prove that $\beta - 1 \le \alpha\cdot\beta$.

(iii) Deduce that $1-\beta^{-1} \le \alpha$ and use this fact, together with the assumptions that $\alpha < 1/2$ and $\beta > 2$, to derive a contradiction.

5. Here is Paul Erdös' (1913–1996) celebrated proof [**64**]. Assume that $\sum 1/p$ converges. Then we can choose an N such that $\sum_{p>N} 1/p < 1/2$; derive a contradiction as follows.

(i) For any $x \in \mathbb{N}$, let A_x be the set of all integers $1 \le n \le x$ such that $n = 1$ or all the prime factors of n are $\le N$; that is, $n = p_1 \cdots p_k$ where the p_j's are prime and $p_j \le N$. Given $n \in A_x$, we can write $n = k^2 m$ where m is square free. Prove that $k \le \sqrt{x}$. From this, deduce that

$$\# A_x \le C\sqrt{x},$$

where $\# A_x$ denotes the number of elements in the set A_x and C is a constant (you can take C to equal the number of square free integers $m \le N$).

(ii) Given $x \in \mathbb{N}$ and a prime p, prove that the number of integers $1 \le n \le x$ divisible by p is no more than x/p.

(iii) Given $x \in \mathbb{N}$, prove that $x - \# A_x$ equals the number of integers $1 \le n \le x$ that are divisible by some prime $p > N$. From this fact and Part (b) together with our assumption that $\sum_{p>N} 1/p < 1/2$, prove that

$$x - \# A_x < \frac{x}{2}.$$

(iv) Using (c) and the inequality $\# A_x \le C\sqrt{x}$ you proved in Part (a), conclude that for any $x \in \mathbb{N}$, we have

$$\sqrt{x} \le 2C.$$

From this derive a contradiction.

6.8. Composition of power series and Bernoulli and Euler numbers

We've kept you in suspense long enough concerning the extraordinary Bernoulli and Euler numbers, so in this section we finally get to these fascinating numbers.

6.8.1. Composition and division of power series.

The Bernoulli and Euler numbers come up when dividing power series, so before we do anything, we need to understand division of power series, and to understand this we first need to consider the composition of power series. The following theorem basically says that the composition of power series is again a power series.

THEOREM 6.33 (**Power series composition theorem**). *If $f(z)$ and $g(z)$ are power series, then the composition $f(g(z))$ can be written as a power series that is valid for all $z \in \mathbb{C}$ such that*

$$\sum_{n=0}^{\infty}|a_n z^n| < \textit{the radius of convergence of } f,$$

where $g(z) = \sum_{n=0}^\infty a_n z^n$.

Proof. Let $f(z) = \sum_{n=0}^\infty b_n z^n$ have radius of convergence R and let $g(z) = \sum_{n=0}^\infty a_n z^n$ have radius of convergence r. Then by Cauchy or Mertens' multiplication theorem, for each m, we can write $g(z)^m$ as a power series:

$$g(z)^m = \left(\sum_{n=0}^\infty a_n\, z^n \right)^m = \sum_{n=0}^\infty a_{mn}\, z^n, \qquad |z| < r.$$

Thus,

$$f(g(z)) = \sum_{m=0}^\infty b_m\, g(z)^m = \sum_{m=0}^\infty \sum_{n=0}^\infty b_m a_{mn}\, z^n.$$

If we are allowed to interchange the order of summation in $f(g(z))$, then our result is proved:

$$f(g(z)) = \sum_{n=0}^\infty \sum_{m=0}^\infty b_m a_{mn}\, z^n = \sum_{n=0}^\infty c_n\, z^n, \qquad \text{where} \ \ c_n = \sum_{m=0}^\infty b_m a_{mn}.$$

Thus, we can focus on interchanging the order of summation in $f(g(z))$. Assume henceforth that

$$\xi := \sum_{n=0}^\infty |a_n z^n| = \sum_{n=0}^\infty |a_n|\, |z|^n < R = \ \text{the radius of convergence of } f;$$

in particular, since $f(\xi) = \sum_{m=0}^\infty b_m \xi^m$ is absolutely convergent,

$$\sum_{m=0}^\infty |b_m|\, \xi^m < \infty. \tag{6.44}$$

Now according to Cauchy's double series theorem, we can interchange the order of summation as long as we can show that

$$\sum_{m=0}^\infty \sum_{n=0}^\infty \left| b_m a_{mn} z^n \right| = \sum_{m=0}^\infty \sum_{n=0}^\infty |b_m|\, |a_{mn}|\, |z|^n < \infty.$$

To prove this, we first claim that the inner summation satisfies the inequality

$$\sum_{n=0}^\infty |a_{mn}|\, |z|^n \le \xi^m. \tag{6.45}$$

To see this, consider the case $m = 2$. Recall that the coefficients a_{2n} are defined via the Cauchy product:

$$g(z)^2 = \left(\sum_{n=0}^\infty a_n\, z^n \right)^2 = \sum_{n=0}^\infty a_{2n}\, z^n \quad \text{where} \quad a_{2n} = \sum_{k=0}^n a_k a_{n-k}.$$

Thus, $|a_{2n}| \le \sum_{k=0}^n |a_k|\, |a_{n-k}|$. On the other hand, we can express ξ^2 via the Cauchy product:

$$\xi^2 = \left(\sum_{n=0}^\infty |a_n|\, |z|^n \right)^2 = \sum_{n=0}^\infty \alpha_n\, |z|^n \quad \text{where} \quad \alpha_n = \sum_{k=0}^n |a_k|\, |a_{n-k}|.$$

Hence, $|a_{2n}| \le \sum_{k=0}^n |a_k|\, |a_{n-k}| = \alpha_n$, so

$$\sum_{n=0}^\infty |a_{2n}|\, |z|^n \le \sum_{n=0}^\infty \alpha_n\, |z|^n = \xi^2,$$

which proves (6.45) for $m = 2$. An induction argument shows that (6.45) holds for all m. Finally, using (6.45) and (6.44) we see that

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left| b_m a_{mn} z^n \right| = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} |b_m| \, |a_{mn}| \, |z|^n \le \sum_{m=0}^{\infty} |b_m| \, \xi^m < \infty,$$

which shows that we can interchange the order of summation in $f(g(z))$ and completes our proof. □

We already know (by Mertens' multiplication theorem for instance) that the product of two power series is again a power series. As a consequence of the following theorem, we get the same statement for division.

THEOREM 6.34 (**Power series division theorem**). *If $f(z)$ and $g(z)$ are power series with positive radii of convergence and with $g(0) \neq 0$, then $f(z)/g(z)$ is also a power series with positive radius of convergence.*

PROOF. Since $f(z)/g(z) = f(z) \cdot (1/g(z))$ and we know that the product of two power series is a power series, all we have to do is show that $1/g(z)$ is a power series. To this end, let $g(z) = \sum_{n=0}^{\infty} a_n \, z^n$ and define

$$\tilde{g}(z) := \frac{1}{a_0} g(z) - 1 = \sum_{n=1}^{\infty} \alpha_n \, z^n,$$

where $\alpha_n = \frac{a_n}{a_0}$ and where we recall that $a_0 = g(0) \neq 0$. Then $\tilde{g}$ has a positive radius of convergence and $\tilde{g}(0) = 0$. Now let $h(z) := \frac{1}{a_0(1+z)}$, which can be written as a geometric series with radius of convergence 1. Note that for $|z|$ small, $\sum_{n=1}^{\infty} |\alpha_n| \, |z|^n < 1$ (why?), thus by the previous theorem, for such z,

$$\frac{1}{g(z)} = \frac{1}{a_0(\tilde{g}(z) + 1)} = h(\tilde{g}(z))$$

has a power series expansion with a positive radius of convergence. □

6.8.2. Bernoulli numbers. See [**120**], [**54**], [**206**], or [**85**] for more information on Bernoulli numbers. Since

$$\frac{e^z - 1}{z} = \frac{1}{z} \cdot \sum_{n=1}^{\infty} \frac{1}{n!} z^n = \sum_{n=1}^{\infty} \frac{1}{n!} z^{n-1} = \sum_{n=0}^{\infty} \frac{1}{(n+1)!} z^n$$

has a power series expansion and equals 1 at $z = 0$, by our division of power series theorem, the quotient $1/((e^z - 1)/z) = z/(e^z - 1)$ also has a power series expansion near $z = 0$. It is customary to denote its coefficients by $B_n/n!$, in which case we can write

$$\boxed{\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} \, z^n} \tag{6.46}$$

where the series has a positive radius of convergence. The numbers B_n are called the **Bernoulli numbers** after Jacob (Jacques) Bernoulli (1654–1705) who discovered them while searching for formulas involving powers of integers; see Problems 3 and 4. We can find a remarkable symbolic equation for these Bernoulli numbers as follows. First, we multiply both sides of (6.46) by $(e^z - 1)/z$ and use Mertens' multiplication theorem to get

$$1 = \left(\sum_{n=0}^\infty \frac{B_n}{n!}\, z^n \right) \cdot \left(\sum_{n=0}^\infty \frac{1}{(n+1)!} z^n \right) = \sum_{n=0}^\infty \sum_{k=0}^n \left(\frac{B_k}{k!} \cdot \frac{1}{(n-k+1)!} \right) z^n.$$

By the identity theorem, the $n = 0$ term on the right must equal 1 while all other terms must vanish. The $n = 0$ term on the right is just B_0, so $B_0 = 1$, and for $n > 1$, we must have $\sum_{k=0}^n \frac{B_k}{k!} \cdot \frac{1}{(n+1-k)!} = 0$. Multiplying this by $(n+1)!$ we get

$$0 = \sum_{k=0}^n \frac{B_k}{k!} \cdot \frac{(n+1)!}{(n+1-k)!} = \sum_{k=0}^n \frac{(n+1)!}{k!(n+1-k)!} \cdot B_k = \sum_{k=0}^n \binom{n+1}{k} B_k,$$

and adding $B_{n+1} = \binom{n+1}{n+1} B_{n+1}$ to both sides of this equation, we get

$$B_{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} B_k.$$

The right-hand side might look familiar from the binomial formula. Recall from the binomial formula that for any complex number a, we have

$$(a+1)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k \cdot 1^{n-k} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k.$$

Notice that the right-hand side of this expression is exactly the right-hand side of the previous equation if put $a = B$ and we make the superscript k into a subscript k. Thus, if we use the notation $\doteqdot$ to mean "equals after making superscripts into subscripts", then we can write

$$\boxed{B^{n+1} \doteqdot (B+1)^{n+1} \quad , \quad n = 1, 2, 3, \ldots \quad \text{with } B_0 = 1.} \tag{6.47}$$

Using the identity (6.47), one can in principle find all the Bernoulli numbers: When $n = 1$, we see that

$$B^2 \doteqdot (B+1)^2 = B^2 + 2B^1 + 1 \quad \Longrightarrow \quad 0 = 2B_1 + 1 \quad \Longrightarrow \quad B_1 = -\frac{1}{2}.$$

When $n = 2$, we see that

$$B^3 \doteqdot (B+1)^3 = B^3 + 3B^2 + 3B^1 + 1 \Longrightarrow 0 = 3B_2 + 3B_1 + 1 \Longrightarrow B_2 = \frac{1}{6}.$$

Here is a partial list through B_{14}:

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_3 = 0,$$
$$B_4 = -\frac{1}{30}, \quad B_5 = B_7 = B_9 = B_{11} = B_{13} = B_{15} = 0,$$
$$B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}, \quad B_{10} = \frac{5}{66}, \quad B_{12} = -\frac{691}{2730}, \quad B_{14} = \frac{7}{6}.$$

These numbers are rational, but besides this fact, there is no known regular pattern these numbers conform to. However, we can easily deduce that all odd Bernoulli numbers greater than one are zero. Indeed, we can rewrite (6.46) as

$$\frac{z}{e^z - 1} + \frac{z}{2} = 1 + \sum_{n=2}^\infty \frac{B_n}{n!}\, z^n. \tag{6.48}$$

The fractions on the left-hand side can be combined into one fraction

$$\frac{z}{e^z-1}+\frac{z}{2}=\frac{z(e^z+1)}{2(e^z-1)}=\frac{z(e^{z/2}+e^{-z/2})}{2(e^{z/2}-e^{-z/2})}, \tag{6.49}$$

which an even function of z. Thus, (see Exercise 1 in Section 6.4)

$$B_{2n+1}=0, \qquad n=1,2,3,\ldots. \tag{6.50}$$

Other properties are given in the exercises (see e.g. Problem 3).

6.8.3. Trigonometric functions. We already know the power series expansions for $\sin z$ and $\cos z$. It turns out that the power series expansions of the other trigonometric functions involve Bernoulli numbers! For example, to find the expansion for $\cot z$, we replace z with $2iz$ in (6.48) and (6.49) to get

$$\frac{iz(e^{iz}+e^{-iz})}{(e^{iz}-e^{-iz})}=1+\sum_{n=2}^{\infty}\frac{B_n}{n!}\,(2iz)^n=1+\sum_{n=1}^{\infty}\frac{B_{2n}}{(2n)!}\,(-1)^n(2z)^{2n},$$

where used that $B_3, B_5, B_7, \ldots$ all vanish in order to sum only over all even Bernoulli numbers. Since $\cot z=\cos z/\sin z$, using the definition of $\cos z$ and $\sin z$ in terms of $e^{\pm iz}$, we see that the left-hand side is exactly $z\cot z$. Thus, we have derived the formula

$$\boxed{z\cot z=\sum_{n=0}^{\infty}(-1)^n\frac{2^{2n}\,B_{2n}}{(2n)!}\,z^{2n}\ .}$$

From this formula, we can get the expansion for $\tan z$ by using the identity

$$2\cot(2z)=2\frac{\cos 2z}{\sin 2z}=2\frac{\cos^2 z-\sin^2 z}{2\sin z\,\cos z}=\cot z-\tan z.$$

Hence,

$$\tan z=\cot z-2\cot(2z)=\sum_{n=0}^{\infty}(-1)^n\frac{2^{2n}\,B_{2n}}{(2n)!}\,z^{2n}-2\sum_{n=0}^{\infty}(-1)^n\frac{2^{2n}\,B_{2n}}{(2n)!}\,2^{2n}z^{2n},$$

which, after combining the terms on the right, takes the form

$$\boxed{\tan z=\sum_{n=1}^{\infty}(-1)^{n-1}\frac{2^{2n}(2^{2n}-1)\,B_{2n}}{(2n)!}\,z^{2n-1}\ .}$$

In Problem 1, we derive the power series expansion of $\csc z$. In conclusion we have power series expansions for $\sin z, \cos z, \tan z, \cot z, \csc z$. What about $\sec z$?

6.8.4. The Euler numbers. It turns out that the expansion for $\sec z$ involves the Euler numbers, which are defined in a similar way as the Bernoulli numbers. By the division of power series theorem, the function $2e^z/(e^{2z}+1)$ has a power series expansion near zero. It is customary to denote its coefficients by $E_n/n!$, so

$$\boxed{\frac{2e^z}{e^{2z}+1}=\sum_{n=0}^{\infty}\frac{E_n}{n!}\,z^n} \tag{6.51}$$

where the series has a positive radius of convergence. The numbers E_n are called the **Euler numbers**. We can get the missing expansion for $\sec z$ as follows. First, observe that

$$\sum_{n=0}^{\infty} \frac{E_n}{n!}\, z^n = \frac{2e^z}{e^{2z}+1} = \frac{2}{e^z + e^{-z}} = \frac{1}{\cosh z} = \operatorname{sech} z,$$

where $\operatorname{sech} z := 1/\cosh z$ is the hyperbolic secant. Since $\operatorname{sech} z$ is an even function (that is, $\operatorname{sech}(-z) = \operatorname{sech} z$) it follows that all E_n with n odd vanish. Hence,

$$\boxed{\operatorname{sech} z = \sum_{n=0}^{\infty} \frac{E_{2n}}{(2n)!}\, z^{2n}\ .} \tag{6.52}$$

In particular, putting iz for z in (6.52) and using that $\cosh(iz) = \cos z$, we get the missing expansion for $\sec z$:

$$\boxed{\sec z = \sum_{n=0}^{\infty} (-1)^n \frac{E_{2n}}{(2n)!}\, z^{2n}\ .}$$

Just as with the Bernoulli numbers, we can derive a symbolic equation for the Euler numbers. To do so, we multiply (6.52) by $\cosh z = \sum_{n=0}^{\infty} \frac{1}{(2n)!} z^{2n}$ and use Mertens' multiplication theorem to get

$$1 = \left(\sum_{n=0}^{\infty} \frac{E_{2n}}{(2n)!}\, z^{2n}\right) \cdot \left(\sum_{n=0}^{\infty} \frac{1}{(2n)!} z^{2n}\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \left(\frac{E_{2k}}{(2k)!} \cdot \frac{1}{(2n-2k)!}\right) z^{2n}.$$

By the identity theorem, the $n = 0$ term on the right must equal 1 while all other terms must vanish. The $n = 0$ term on the right is just E_0, so $E_0 = 1$, and for $n > 1$, we must have $\sum_{k=0}^{n} \frac{E_{2k}}{(2k)!} \cdot \frac{1}{(2n-2k)!} = 0$. Multiplying this by $(2n)!$ we get

$$0 = \sum_{k=0}^{n} \frac{E_{2k}}{(2k)!} \cdot \frac{(2n)!}{(2n-2k)!} = \sum_{k=0}^{n} \frac{(2n)!}{(2k)!(2n-2k)!} \cdot E_{2k}. \tag{6.53}$$

Now from the binomial formula, for any complex number a, we have

$$\begin{aligned}(a+1)^{2n} + (a-1)^{2n} &= \sum_{k=0}^{2n} \frac{(2n)!}{k!(2n-k)!} a^k + \sum_{k=0}^{2n} \frac{(2n)!}{k!(2n-k)!} a^k (-1)^{2n-k} \\ &= \sum_{k=0}^{2n} \frac{(2n)!}{k!(2n-k)!} a^k + \sum_{k=0}^{2n} \frac{(2n)!}{k!(2n-k)!} a^k (-1)^k \\ &= \sum_{k=0}^{2n} \frac{(2n)!}{(2k)!(2n-2k)!} a^{2k},\end{aligned}$$

since all the odd terms cancel. Notice that the right-hand side of this expression is exactly the right-hand side of (6.53) if put $a = E$ and we make the superscript $2k$ into a subscript $2k$. Thus,

$$\boxed{(E+1)^{2n} + (E-1)^{2n} \doteq 0\ \ ,\ \ n = 1, 2, \ldots \ \text{ with } E_0 = 1 \text{ and } E_{\text{odd}} = 0.} \tag{6.54}$$

Using the identity (6.54), one can in principle find all the Euler numbers: When $n = 1$, we see that

$$(E^2 + 2E^1 + 1) + (E^2 - 2E^1 + 1) \doteq 0 \quad\Longrightarrow\quad 2E_2 + 2 = 0 \quad\Longrightarrow\quad E_2 = -1.$$

Here is a partial list through E_{12}:

$$E_0 = 1, \quad E_1 = E_2 = E_3 = \cdots = 0 \ \ (E_{\text{odd}} = 0), \quad E_2 = -1, \quad E_4 = 5$$
$$E_6 = -61, \quad E_8 = 1385, \quad E_{10} = -50,521, \quad E_{12} = 2,702,765, \quad \ldots.$$

These numbers are all integers, but besides this fact, there is no known regular pattern these numbers conform to.

Exercises 6.8.

1. Recall that $\csc z = 1/\sin z$. Prove that $\csc z = \cot z + \tan(z/2)$, and from this identity deduce that

$$\boxed{z \csc z = \sum_{n=0}^{\infty} (-1)^{n-1} \frac{(2^{2n} - 2)\, B_{2n}}{(2n)!}\, z^{2n}.}$$

2. (a) Let $f(z) = \sum a_n z^n$ and $g(z) = \sum b_n z^n$ with $b_0 \neq 0$ be power series with positive radii of convergence. Show that $f(z)/g(z) = \sum c_n z^n$ where $\{c_n\}$ is the sequence defined recursively as follows:

$$c_0 = \frac{a_0}{b_0} \quad , \quad b_0 c_n = a_n - \sum_{k=1}^{n} b_k\, c_{n-k}.$$

(b) Use Part (a) to find the first few coefficients of the expansion for $\tan z = \sin z/\cos z$.

3. (Cf. [**120**, p. 526] which is reproduced in [**166**]) In this and the next problem we give an elegant application of the theory of Bernoulli numbers to determine the sum of the first k-th powers of integers, Bernoulli's original motivation for his numbers.

(i) For $n \in \mathbb{N}$, derive the formula

$$1 + e^z + e^{2z} + \cdots + e^{nz} = \frac{z}{e^z - 1} \cdot \frac{e^{(n+1)z} - 1}{z}.$$

(ii) Writing each side of this identity as a power series (on the right, you need to use the Cauchy product), derive the formula

$$\boxed{1^k + 2^k + \cdots + n^k = \sum_{j=0}^{k} \binom{k}{j} B_j\, \frac{(n+1)^{k+1-j}}{k+1-j}, \qquad k = 1, 2, \ldots.} \tag{6.55}$$

Plug in $k = 1, 2, 3$ to derive some pretty formulas!

4. Here's another proof of (6.55) that is aesthetically appealing.

(i) Prove that for a complex number a and natural numbers k, n,

$$(n + 1 + a)^{k+1} - (n + a)^{k+1} = \sum_{j=1}^{k+1} \binom{k+1}{j} n^{k+1-j} \Big((a+1)^j - a^j \Big).$$

(ii) Replacing a with B, prove that

$$\boxed{1^k + 2^k + \cdots + n^k \doteqdot \frac{1}{k+1} \big\{ (n + 1 + B)^{k+1} - B^{k+1} \big\}.}$$

Suggestion: Look for a telescoping sum and recall that $(B + 1)^j \doteqdot B^j$ for $j \geq 2$.

5. The n-th Bernoulli polynomial $B_n(t)$ is by definition, $n!$ times the coefficient of z^n in the power series expansion in z of the function $f(z, t) := ze^{zt}/(e^z - 1)$; that is,

$$\boxed{\frac{z\, e^{zt}}{e^z - 1} = \sum_{n=0}^{\infty} \frac{B_n(t)}{n!}\, z^n.} \tag{6.56}$$

(a) Prove that $B_n(t) = \sum_{k=0}^{n} \binom{n}{k} B_k\, t^{n-k}$ where the B_k's are the Bernoulli numbers. Thus, the first few Bernoulli polynomials are

$$B_0(t) = 1, \quad B_1(t) = t - \frac{1}{2}, \quad B_2(t) = t^2 - t + \frac{1}{6}, \quad B_3(t) = t^3 - \frac{3}{2}t^2 + \frac{1}{2}t.$$

(b) Prove that $B_n(0) = B_n$ for $n = 0, 1, \ldots$ and that $B_n(0) = B_n(1) = B_n$ for $n \neq 1$. Suggestion: Show that $f(z, 1) = z + f(z, 0)$.

(c) Prove that $B_n(t+1) - B_n(t) = nt^{n-1}$ for $n = 0, 1, 2, \ldots$. Suggestion: Show that $f(z, t+1) - f(z, t) = ze^{zt}$.

(d) Prove that $B_{2n+1}(0) = 0$ for $n = 1, 2, \ldots$ and $B_{2n+1}(1/2) = 0$ for $n = 0, 1, \ldots$.

6.9. The logarithmic, binomial, arctangent series, and γ

From elementary calculus, you might have seen the logarithmic, binomial, and arctangent series (discovered by Nicolaus Mercator (1620–1687), Sir Isaac Newton (1643–1727), and Madhava of Sangamagramma (1350–1425), respectively):

$$\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n \ , \ (1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n \ , \ \arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$$

where $\alpha \in \mathbb{R}$. (Below we'll discuss the meaning of $\binom{\alpha}{n}$.) I can bet that you used calculus (derivatives and integrals) to derive these formulæ. In this section we'll derive even more general complex versions of these formulæ without derivatives!

6.9.1. The binomial coefficients. From our familiar binomial theorem, we know that for any $z \in \mathbb{C}$ and $k \in \mathbb{N}$, we have $(1+z)^k = \sum_{n=0}^{k} \binom{k}{n} z^n$, where $\binom{k}{0} := 1$ and for $n = 1, 2, \ldots, k$,

$$(6.57) \qquad \binom{k}{n} := \frac{k!}{n!(k-n)!} = \frac{1 \cdot 2 \cdots k}{n! \cdot 1 \cdot 2 \cdots (k-n)} = \frac{k(k-1)\cdots(k-n+1)}{n!}.$$

The formula $(1+z)^k = \sum_{n=0}^{k} \binom{k}{n} z^n$ trivially holds when $k = 0$ too. Another way to express this formula is

$$(1+z)^k = 1 + \sum_{n=1}^{k} \frac{k(k-1)\cdots(k-n+1)}{n!} z^n.$$

With this motivation, given any complex number α, we define the **binomial coefficient** $\binom{\alpha}{n}$ for any nonnegative integer n as follows: $\binom{\alpha}{0} = 1$ and for $n > 0$,

$$(6.58) \qquad \binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}.$$

Note that if $\alpha = 0, 1, 2, \ldots$, then we see that all $\binom{\alpha}{n}$ vanish for $n \geq \alpha + 1$ and $\binom{\alpha}{n}$ is exactly the usual binomial coefficient (6.57). In the following lemma, we derive an identity that will be useful later.

Lemma 6.35. *For any $\alpha, \beta \in \mathbb{C}$, we have*

$$\binom{\alpha+\beta}{n} = \sum_{k=0}^{n} \binom{\alpha}{k} \binom{\beta}{n-k}, \qquad n = 0, 1, 2, \ldots.$$

Proof. Throughout this proof, we put $\mathbb{N}_0 := \{0, 1, 2, 3, \ldots\}$.

Step 1: First of all, our lemma holds when both α and β are in $\mathbb{N}_0$. Indeed, if $\alpha = p, \beta = q$ are in $\mathbb{N}_0$, then expressing both sides of the identity $(1+z)^{p+q} = (1+z)^p(1+z)^q$ using the binomial formula, we obtain

$$\sum_{n=0}^{p+q} \binom{p+q}{n} z^n = \left(\sum_{k=0}^{p} \binom{p}{k} z^k\right) \cdot \left(\sum_{k=0}^{q} \binom{q}{k} z^k\right)$$
$$= \sum_{n=0}^{p+q} \left(\sum_{k=0}^{n} \binom{p}{k} \binom{q}{n-k}\right) z^n,$$

where at the last step we formed the Cauchy product of $(1+z)^p(1+z)^q$. By the identity theorem we must have

$$\binom{p+q}{n} = \sum_{k=0}^{n} \binom{p}{k} \binom{q}{n-k}, \quad \text{for all } p, q, n \in \mathbb{N}_0.$$

Step 2: Assume now that $\beta = q \in \mathbb{N}_0$, $n \in \mathbb{N}_0$, and define $f : \mathbb{C} \longrightarrow \mathbb{C}$ by

$$f(z) := \binom{z+q}{n} - \sum_{k=0}^{n} \binom{z}{k} \binom{q}{n-k}.$$

In view of the definition (6.58) of the binomial coefficient, it follows that $f(z)$ is a polynomial in z of degree at most n. Moreover, by **Step 1** we know that $f(p) = 0$ for all $p \in \mathbb{N}_0$. In particular, the polynomial $f(z)$ has more than n roots. Therefore, $f(z)$ must be the zero polynomial, so in particular, given any $\alpha \in \mathbb{C}$, we have $f(\alpha) = 0$; that is,

$$\binom{\alpha+q}{n} = \sum_{k=0}^{n} \binom{\alpha}{k} \binom{q}{n-k}, \quad \text{for all } \ \alpha \in \mathbb{C}\,, \ q, n \in \mathbb{N}_0.$$

Step 3: Let $\alpha \in \mathbb{C}$, $n \in \mathbb{N}_0$, and define $g : \mathbb{C} \longrightarrow \mathbb{C}$ by

$$g(z) := \binom{\alpha+z}{n} - \sum_{k=0}^{n} \binom{\alpha}{k} \binom{z}{n-k}.$$

As with the function $f(z)$ in **Step 2**, $g(z)$ is a polynomial in z of degree at most n. Also, by **Step 2** we know that $g(q) = 0$ for all $q \in \mathbb{N}_0$ and consequently, $g(z)$ must be the zero polynomial. In particular, given any $\beta \in \mathbb{C}$, we have $g(\beta) = 0$, which completes our proof. □

6.9.2. The complex logarithm and binomial series. In Theorem 6.37 we shall derive (along with a power series for Log) the **binomial series**:

$$(1+z)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n = 1 + \alpha\, z + \frac{\alpha(\alpha-1)}{1!}\, z^2 + \cdots, \quad |z| < 1. \tag{6.59}$$

Let us define $f(\alpha, z) := \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n$. Our goal is to show that $f(\alpha, z) = (1+z)^\alpha$ for all $\alpha \in \mathbb{C}$ and $|z| < 1$, where

$$(1+z)^\alpha := \exp(\alpha \operatorname{Log}(1+z))$$

with Log the principal logarithm of the complex number $1+z$. If $\alpha = k = 0, 1, 2, \ldots$, then we already know that all the $\binom{k}{n}$ vanish for $n \geq k+1$ and these binomial coefficients are the usual ones, so $f(k, z)$ converges with sum $f(k, z) = (1+z)^k$. To see that $f(\alpha, z)$ converges for all other α, assume that $\alpha \in \mathbb{C}$ is not a nonnegative integer. Then setting $a_n = \binom{\alpha}{n}$, we have

$$\left|\frac{a_n}{a_{n+1}}\right| = \left|\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} \cdot \frac{(n+1)!}{\alpha(\alpha-1)\cdots(\alpha-n)}\right| = \frac{n+1}{|\alpha-n|},$$

which approaches 1 as $n \to \infty$. Thus, the radius of convergence of $f(\alpha, z)$ is 1 (see (6.12)). In conclusion, $f(\alpha, z)$ is convergent for all $\alpha \in \mathbb{C}$ and $|z| < 1$.

We now prove the real versions of the logarithm series and the binomial series (6.59); see Theorem 6.37 below for the more general complex version. It is worth emphasizing that we do not use the advanced technology of the differential and integral calculus to derive these formulas!

Lemma 6.36. *For all $x \in \mathbb{R}$ with $|x| < 1$, we have*

$$\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$$

and for all $\alpha \in \mathbb{C}$ and $x \in \mathbb{R}$ with $|x| < 1$, we have

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n = 1 + \alpha\, x + \frac{\alpha(\alpha-1)}{1!}\, x^2 + \cdots .$$

Proof. We prove this lemma in three steps.

Step 1: We first show that $f(r, x) = (1+x)^r$ for all $r = p/q \in \mathbb{Q}$ where $p, q \in \mathbb{N}$ with q *odd* and $x \in \mathbb{R}$ with $|x| < 1$. To see this, observe for any $z \in \mathbb{C}$ with $|z| < 1$, taking the Cauchy product of $f(\alpha, z)$ and $f(\beta, z)$ and using our lemma, we obtain

$$f(\alpha, z) \cdot f(\beta, z) = \sum_{n=0}^{\infty} \left(\sum_{j=0}^{n} \binom{\alpha}{j} \binom{\beta}{n-j} \right) z^n = \sum_{n=0}^{\infty} \binom{\alpha+\beta}{n} z^n = f(\alpha+\beta, z).$$

By induction it easily follows that

$$f(\alpha_1, z) \cdot f(\alpha_2, z) \cdots f(\alpha_k, z) = f(\alpha_1 + \alpha_2 + \cdots + \alpha_k, z).$$

Using this identity, we obtain

$$f(1/q, z)^q = \underbrace{f(1/q, z) \cdots f(1/q, z)}_{q \text{ times}} = f(\underbrace{1/q + \cdots + 1/q}_{q \text{ times}}, z) = f(1, z) = 1 + z.$$

Now put $z = x \in \mathbb{R}$ with $|x| < 1$ and let $q \in \mathbb{N}$ be odd. Then $f(1/q, x)^q = 1 + x$, so taking q-th roots, we get $f(1/q, x) = (1+x)^{1/q}$. Here we used that every real number has a unique q-th root, which holds because q is odd — for q even we could only conclude that $f(1/q, x) = \pm(1+x)^{1/q}$ (unless we checked that $f(1/q, x)$ is positive, then we would get $f(1/q, x) = (1+x)^{1/q}$). Therefore,

$$\begin{aligned} f(r, x) = f(p/q, x) = f(\underbrace{1/q + \cdots + 1/q}_{p \text{ times}}, x) &= \underbrace{f(1/q, x) \cdots f(1/q, x)}_{p \text{ times}} \\ &= f(1/q, x)^p = (1+x)^{p/q} = (1+x)^r. \end{aligned}$$

Step 2: Second, we prove that for any given $z \in \mathbb{C}$ with $|z| < 1$, $f(\alpha, z)$ can be written as a power series in α that converges for all $\alpha \in \mathbb{C}$:

$$f(\alpha, z) = 1 + \sum_{m=1}^{\infty} a_m(z)\, \alpha^m;$$

in particular, since we know that power series are continuous, $f(\alpha, z)$ is a continuous function of $\alpha \in \mathbb{C}$. Here, the coefficients $a_m(z)$ depend on z (which we'll see are power series in z) and we'll show that

$$a_1(z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n. \tag{6.60}$$

To prove these statements, note that for $n \geq 1$, $\alpha(\alpha-1)\cdots(\alpha-n+1)$ is a polynomial of degree n in α, so for $n \geq 1$ we can write

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} = \sum_{m=1}^{n} a_{mn}\, \alpha^m, \tag{6.61}$$

for some coefficients a_{mn}. Defining $a_{mn} = 0$ for $m = n+1, n+2, n+3, \ldots$, we can write $\binom{\alpha}{n} = \sum_{m=0}^{\infty} a_{mn}\, \alpha^m$. Hence,

$$f(z, \alpha) = 1 + \sum_{n=1}^{\infty} \binom{\alpha}{n} z^n = 1 + \sum_{n=1}^{\infty} \Bigg(\sum_{m=1}^{\infty} a_{mn}\, \alpha^m \Bigg) z^n. \tag{6.62}$$

To make this a power series in α, we need to switch the order of summation, which we can do by Cauchy's double series theorem if we can demonstrate absolute convergence by showing that

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \big|a_{mn}\, \alpha^m z^n\big| = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |a_{mn}|\, |\alpha|^m\, |z|^n < \infty.$$

To verify this, we first observe that for all $\alpha \in \mathbb{C}$, we have

$$\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{n!} = \sum_{m=1}^{n} b_{mn}\, \alpha^m, \tag{6.63}$$

where the b_{mn}'s are nonnegative real numbers. (This is certainly plausible because the numbers $1, 2, \ldots, n-1$ on the left each come with positive signs; any case, this statement can be verified by induction for instance.) We secondly observe that replacing α with $-\alpha$ in (6.61), we get

$$\begin{aligned} \sum_{m=1}^{n} a_{mn}\, (-1)^m \alpha^m &= \frac{-\alpha(-\alpha-1)\cdots(-\alpha-n+1)}{n!} \\ &= (-1)^n \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)}{n!} = \sum_{m=1}^{n} (-1)^n b_{mn}\, \alpha^m. \end{aligned}$$

By the identity theorem, we have $a_{mn}(-1)^m = (-1)^n b_{mn}$. In particular, $|a_{mn}| = b_{mn}$ since $b_{mn} > 0$, therefore in view of (6.63), we see that

$$\sum_{m=0}^{\infty} |a_{mn}|\, |\alpha|^m = \sum_{m=0}^{n} |a_{mn}|\, |\alpha|^m = \sum_{m=0}^{n} b_{mn}\, |\alpha|^m = \frac{|\alpha|(|\alpha|+1)\cdots(|\alpha|+n-1)}{n!}.$$

Therefore,

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |a_{mn}|\, |\alpha|^m\, |z|^n = \sum_{n=1}^{\infty} \frac{|\alpha|(|\alpha|+1)\cdots(|\alpha|+n-1)}{n!}\, |z|^n.$$

Using the now very familiar ratio test it's easily checked that, since $|z| < 1$, the series on the right converges. Thus, we can iterate sums in (6.62) and conclude that

$$f(\alpha, z) = 1 + \sum_{n=1}^{\infty} \bigg(\sum_{m=1}^{\infty} a_{mn}\, \alpha^m \bigg) z^n = 1 + \sum_{m=1}^{\infty} \bigg(\sum_{n=1}^{\infty} a_{mn}\, z^n \bigg) \alpha^m .$$

Thus, $f(\alpha, z)$ is indeed a power series in α. To prove (6.60), we just need to determine the coefficient of α^1 in (6.61), which we see is just

$$\begin{aligned} a_{1n} &= \text{coefficient of } \alpha \text{ in } \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!} \\ &= \frac{(-1)(-2)(-3)\cdots(-n+1)}{n!} = (-1)^{n-1}\frac{(n-1)!}{n!} = \frac{(-1)^{n-1}}{n}. \end{aligned}$$

Therefore,

$$a_1(z) = \sum_{n=1}^{\infty} a_{1n}\, z^n = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n ,$$

just as we stated in (6.60). This completes **Step 2**.

Step 3: We are finally ready to prove our theorem. Let $x \in \mathbb{R}$ with $|x| < 1$. By **Step 2**, we know that for any $\alpha \in \mathbb{C}$,

$$f(\alpha, x) = 1 + \sum_{m=1}^{\infty} a_m(x)\, \alpha^m$$

is a power series in α. However,

$$(1+x)^\alpha = \exp(\alpha \log(1+x)) = \sum_{n=0}^{\infty} \frac{1}{n!} \log(1+x)^n \cdot \alpha^n$$

is also a power series in $\alpha \in \mathbb{C}$. By **Step 1**, $f(\alpha, x) = (1+x)^\alpha$ for all $\alpha \in \mathbb{Q}$ with $\alpha > 0$ having odd denominators. The identity theorem applies to this situation (why?), so we must have $f(\alpha, x) = (1+x)^\alpha$ for all $\alpha \in \mathbb{C}$. Also by the identity theorem, the coefficients of α^n must be identical; in particular, the coefficients of α^1 are identical: $a_1(x) = \log(1+x)$. Now (6.60) implies the series for $\log(1+x)$. □

Using this lemma and the identity theorem, we are ready to generalize these formulas for real x to formulas for complex z.

THEOREM 6.37 (**The complex logarithm and binomial series**). *We have*

$$\boxed{\operatorname{Log}(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n, \quad |z| \le 1,\ z \ne -1,}$$

and given any $\alpha \in \mathbb{C}$, we have

$$\boxed{(1+z)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n = 1 + \alpha\, z + \frac{\alpha(\alpha-1)}{1!}\, z^2 + \cdots , \quad |z| < 1.}$$

PROOF. We prove this theorem first for $\operatorname{Log}(1+z)$, then for $(1+z)^\alpha$.

Step 1: Let us define $f(z) := \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n$. Then one can check that the radius of convergence of $f(z)$ is 1, so by our power series composition theorem, we know that $\exp(f(z))$ can be written as a power series:

$$\exp(f(z)) = \sum_{n=0}^{\infty} a_n\, z^n, \qquad |z| < 1.$$

Restricting to real values of z, by our lemma we know that $f(x) = \log(1 + x)$, so

$$\sum_{n=0}^{\infty} a_n\, x^n = \exp(f(x)) = \exp(\log(1 + x)) = 1 + x.$$

By the identity theorem for power series, we must have $a_0 = 1, a_1 = 1$, and all other $a_n = 0$. Thus, $\exp(f(z)) = 1 + z$. Since $\exp(\mathrm{Log}(1 + z)) = 1 + z$ as well, we have

$$\exp(f(z)) = \exp(\mathrm{Log}(1 + z)),$$

which implies that $f(z) = \mathrm{Log}(1 + z) + 2\pi i k$ for some integer k. Setting $z = 0$ shows that $k = 0$ and hence proves that $\mathrm{Log}(1 + z) = f(z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n$.

We now prove that $\mathrm{Log}(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n$ holds for $|z| = 1$ with $z \neq -1$ (note that for $z = -1$, both sides of this equality are not defined). If $|z| = 1$, then we can write $z = -e^{ix}$ with $x \in (0, 2\pi)$. Recall from Example 6.4 in Section 6.1 that for any $x \in (0, 2\pi)$, the series $\sum_{n=1}^{\infty} \frac{e^{inx}}{n}$ converges. Hence, as

$$-\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n = \sum_{n=1}^{\infty} \frac{(-1)^n(-e^{ix})^n}{n} = \sum_{n=1}^{\infty} \frac{e^{inx}}{n}, \tag{6.64}$$

it follows that $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n$ converges for $|z| = 1$ with $z \neq -1$. Now fix a point z_0 with $|z_0| = 1$ and $z_0 \neq -1$, and let us take $z \to z_0$ through the straight line from $z = 0$ to $z = z_0$ (that is, let $z = tz_0$ where $0 \leq t \leq 1$ and take $t \to 1^-$). Since the ratio

$$\frac{|z_0 - z|}{1 - |z|} = \frac{|z_0 - tz_0|}{1 - |tz_0|} = \frac{|z_0 - tz_0|}{1 - t} = \frac{|1 - t|}{1 - t} = 1,$$

which bounded by a fixed constant, by Abel's theorem (Theorem 6.20), it follows that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z_0^n = \lim_{z \to z_0} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n = \lim_{z \to z_0} \mathrm{Log}(1 + z) = \mathrm{Log}(1 + z_0),$$

where we used that $\mathrm{Log}(1 + z)$ is continuous.

Step 2: Let $\alpha \in \mathbb{C}$. To prove the binomial series, we note that by the power series composition theorem, $(1 + z)^\alpha = \exp(\alpha\, \mathrm{Log}(1 + z))$, being the composition of exp and Log, can be written as a power series:

$$(1 + z)^\alpha = \sum_{n=0}^{\infty} b_n\, z^n, \qquad |z| < 1.$$

Restricting to real $z = x \in \mathbb{R}$ with $|x| < 1$, by our lemma we know that $(1 + x)^\alpha = f(\alpha, x)$. Hence, by the identity theorem, we must have $(1 + z)^\alpha = f(\alpha, z)$ for all $z \in \mathbb{C}$ with $|z| < 1$. This proves the binomial series. □

For any $z \in \mathbb{C}$ with $|z| < 1$, we have $\operatorname{Log}\big((1+z)/(1-z)\big) = \operatorname{Log}(1+z) - \operatorname{Log}(1-z)$. Therefore, we can use this theorem to prove that (see Problem 1)

$$\boxed{\frac{1}{2}\operatorname{Log}\left(\frac{1+z}{1-z}\right) = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{2n+1}.} \tag{6.65}$$

Here's another consequence of Theorem 6.37.

Example 6.39. In the proof of Theorem 6.37 we used that, for $x \in (0, 2\pi)$, the series $\sum_{n=1}^{\infty} \frac{e^{inx}}{n} = \sum_{n=1}^{\infty} \frac{\cos nx}{n} + \sum_{n=1}^{\infty} \frac{\sin nx}{n}$ converges. We shall prove that

$$\boxed{\sum_{n=1}^{\infty} \frac{\cos nx}{n} = \log\big(2\sin(x/2)\big) \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{\sin nx}{n} = \frac{x-\pi}{2}.}$$

To see this, recall from (6.64) that, with $z = -e^{ix}$, we have

$$\sum_{n=1}^{\infty} \frac{\cos nx}{n} + i\sum_{n=1}^{\infty} \frac{\sin nx}{n} = -\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n = -\operatorname{Log}(1+z) = -\operatorname{Log}(1-e^{ix}).$$

We can write

$$1 - e^{ix} = e^{ix/2}(e^{-ix/2} - e^{ix/2}) = -2ie^{ix/2}\sin(x/2) = 2\sin(x/2)e^{ix/2 - i\pi/2}.$$

Hence, by definition of Log, we have

$$\operatorname{Log}(1-e^{ix}) = \log\big(2\sin(x/2)\big) + i\frac{x-\pi}{2}.$$

This proves our result.

6.9.3. Gregory-Madhava series and formulæ for γ. Recall from Section 4.9 that

$$\operatorname{Arctan} z = \frac{1}{2i}\operatorname{Log}\frac{1+iz}{1-iz}.$$

Using (6.65), we get the famous formula first discovered by Madhava of Sangamagramma (1350–1425) around 1400 and rediscovered over 200 years later in Europe by James Gregory (1638–1675), who found it in 1671! In fact, the mathematicians of Kerala in southern India not only discovered the arctangent series, they also discoved the infinite series for sine and cosine, but their results were written up in Sanskrit and not brought to Europe until the 1800's. For more history on this fascinating topic, see the articles [**111**], [**190**], and the website [**172**].

THEOREM 6.38. *For any complex number z with $|z| < 1$, we have*

$$\boxed{\operatorname{Arctan} z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{2n+1}, \qquad \textbf{\textit{Gregory-Madhava's series}}.}$$

This series is commonly known as **Gregory's arctangent series**, but we shall call it the **Gregory-Madhava arctangent series** because of Madhava's contribution to this series. Setting $z = x$, a real variable, we obtain the usual formula learned in elementary calculus:

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}.$$

In Problem 5 we prove the following stunning formulæ for the Euler-Mascheroni constant γ in terms of the Riemann ζ-function $\zeta(z)$:

$$\boxed{\begin{aligned}\gamma &= \sum_{n=2}^{\infty} \frac{(-1)^n}{n}\zeta(n)\\ &= 1 - \sum_{n=2}^{\infty} \frac{1}{n}\big(\zeta(n) - 1\big)\\ &= \frac{3}{2} - \log 2 - \sum_{n=2}^{\infty} \frac{(-1)^n}{n}(n-1)\big(\zeta(n)-1\big).\end{aligned}} \tag{6.66}$$

The first two formulas are due to Euler and the last one to Philippe Flajolet and Ilan Vardi (see [**203**, pp. 4,5], [**75**]).

Exercises 6.9.

1. Fill in the details in the proof of formula (6.65).
2. Derive the remarkably pretty formulas:

$$\boxed{2(\operatorname{Arctan} z)^2 = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+2}\Big(1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n+1}\Big) z^{2n+2},}$$

and the formula

$$\boxed{\frac{1}{2}(\operatorname{Log}(1+z))^2 = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+2}\Big(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n+1}\Big) z^{n+2},}$$

both valid for $|z| < 1$.

3. Before looking at the next section, prove that

$$\arctan x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \quad \text{and} \quad \log(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n} x^n$$

are valid for $-1 < x \le 1$. Suggestion: I know you are Abel to do this! From these facts, derive the formulas

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots \quad \text{and} \quad \log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

4. For $\alpha \in \mathbb{R}$, prove that $\sum_{n=0}^{\infty} \binom{\alpha}{n}$ converges if and only if $\alpha \le 0$ or $\alpha \in \mathbb{N}$, in which case,

$$2^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n}.$$

Suggestion: To prove convergence use Gauss' test.

5. Prove the exquisite formulas

$$(a)\ \sum_{n=1}^{\infty} \frac{1}{n}\frac{z^n}{1-z^n} = \sum_{n=1}^{\infty} \operatorname{Log} \frac{1}{1-z^n}, \qquad |z| < 1,$$

$$(b)\ \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\frac{z^n}{1-z^n} = \sum_{n=1}^{\infty} \operatorname{Log}(1+z^n), \qquad |z| < 1.$$

Suggestion: Cauchy's double series theorem.

6. In this problem, we prove the stunning formulæ in (6.66).
   (i) Using the first formula for γ in Problem 7a of Exercises 4.6, prove that $\gamma = \sum_{n=1}^{\infty} f\Big(\frac{1}{n}\Big)$ where $f(z) = \sum_{n=2}^{\infty} \frac{(-1)^n}{n} z^n$.

(ii) Prove that $\gamma = 1 - \log 2 + \sum_{n=2}^\infty \frac{(-1)^n}{n}(\zeta(n) - 1)$ using (i) and Problem 10 in Exercises 6.5. Show that this formula is equivalent to the first formula in (6.66).

(iii) Using the second and third formulas in Problem 7a of Exercises 4.6, derive the second and third formulas in (6.66).

6.10. ★ π, Euler, Fibonacci, Leibniz, Madhava, and Machin

In this section, we continue our fascinating study of formulas for π that we initiated in Section 4.10. In particular, we derive (using a very different method from the one presented in Section 5.2) Gregory-Leibniz-Madhava's formula for $\pi/4$, formulas for π discovered by Euler involving the arctangent function and even the Fibonacci numbers, and finally, we look at Machin's formula for π, versions of which has been used to compute trillions of digits of π by Yasumasa Kanada and his coworkers at the University of Tokyo.[6] For other formulas for $\pi/4$ in terms of arctangents, see the articles [**132, 97**]. For more on the history of computations of π, see [**13**], and for interesting historical facets on π in general, see [**10**], [**47, 48**]. The website [**207**] has tons of information.

6.10.1. Gregory-Leibniz-Madhava's formula for $\pi/4$, Proof II.

Recall Gregory-Madhava's formula for real values:

$$\arctan x = \sum_{n=0}^\infty (-1)^{n-1} \frac{x^{2n-1}}{2n-1}.$$

By the alternating series theorem, we know that $\sum_{n=0}^\infty (-1)^{n-1}/(2n-1)$ converges, therefore by Abel's limit theorem (Theorem 6.20) we know that

$$\frac{\pi}{4} = \lim_{x\to 1-} \arctan x = \sum_{n=0}^\infty (-1)^{n-1} \frac{1}{2n-1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots .$$

Therefore, we obtain another derivation of

$$\boxed{\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots , \qquad \textbf{Gregory-Leibniz-Madhava's series}.}$$

Madhava of Sangamagramma (1350–1425) was the first to discover this formula, over 200 years before James Gregory (1638–1675) or Gottfried Leibniz (1646–1716) were even born! Note that the Gregory-Leibniz-Madhava's series is really just a special case of Gregory-Madhava's formula for $\arctan x$ (just set $x = 1$), which recall was discovered in 1671 by Gregory and again, much earlier by Madhava. Leibniz discovered the formula for $\pi/4$ (using geometric arguments) around 1673. Although there is no published record of Gregory noting the formula for $\pi/4$ (he published few of his mathematical results plus he died at only 37 years old), it would be hard to believe that he didn't know about the formula for $\pi/4$. For more history, including Nilakantha Somayaji's (1444–1544) contribution, see [**190, 172, 111, 38**].

Example 6.40. Let us say that we want to approximate $\pi/4$ by Gregory-Leibniz-Madhava's series to within, say a reasonable amount of 7 decimal places.

[6] *The value of π has engaged the attention of many mathematicians and calculators from the time of Archimedes to the present day, and has been computed from so many different formulae, that a complete account of its calculation would almost amount to a history of mathematics. James Glaisher (1848–1928)* [**82**].

Then denoting the n-th partial sum of Gregory-Leibniz-Madhava's series by s_n, according to the alternating series error estimate, we want

$$\left|\frac{\pi}{4} - s_n\right| \leq \frac{1}{2n+1} < 0.00000005 = 5 \times 10^{-8},$$

which implies that $2n + 1 > 10^8/5$, which works for $n \geq 10,000,000$. Thus, we can approximate $\pi/4$ by the n-th partial sum of Gregory-Leibniz-Madhava's series by taking *ten million* terms! Thus, although Gregory-Leibniz-Madhava's series is beautiful, it is quite useless to compute π.

Example 6.41. From Gregory-Leibniz-Madhava's formula, we can easily derive the breath-taking formula (see Problem 4)

$$\boxed{\pi = \sum_{n=2}^{\infty} \frac{3^n - 1}{4^n}\,\zeta(n+1),} \tag{6.67}$$

due to Philippe Flajolet and Ilan Vardi (see [**204**, p. 1], [**232, 75**]).

6.10.2. Euler's arctangent formula and the Fibonacci numbers. In 1738, Euler derived a very pretty two-angle arctangent expression for π:

$$\boxed{\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{3}.} \tag{6.68}$$

This formula is very easy to derive. We start off with the addition formula for tangent (see (4.34), but now considering real variables)

$$\frac{\tan\theta + \tan\phi}{1 - \tan\theta\,\tan\phi} = \tan(\theta + \phi), \tag{6.69}$$

where it is assumed that $1 - \tan\theta\,\tan\phi \neq 0$. Let $x = \tan\theta$ and $y = \tan\phi$ and assume that $-\pi/2 < \theta + \phi < \pi/2$. Then taking arctangents of both sides of the above equation, we obtain

$$\arctan\left(\frac{x+y}{1-xy}\right) = \theta + \phi,$$

or after putting the left-hand in terms of x, y, we get

$$\arctan\left(\frac{x+y}{1-xy}\right) = \arctan x + \arctan y. \tag{6.70}$$

Setting $x = 1/2$ and $y = 1/3$ and using that

$$\frac{x+y}{1-xy} = \frac{5/6}{1-5/6} = 1,$$

we get

$$\arctan 1 = \arctan\frac{1}{2} + \arctan\frac{1}{3}.$$

This expression is just (6.68).

In Problem 9 of Exercises 2.2 we studied the **Fibonacci sequence**, named after Leonardo Fibonacci (1170–1250): $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$ and you proved that for every n,

$$F_n = \frac{1}{\sqrt{5}}\Big[\Phi^n - (-\Phi)^{-n}\Big], \qquad \Phi = \frac{1+\sqrt{5}}{2}. \tag{6.71}$$

We can use (6.68) and (6.70) to derive the following fascinating formula for $\pi/4$ in terms of the (odd-indexed) Fibonacci numbers due to Lehmer [**131**] (see Problem 2 and [**133**]):

$$\boxed{\frac{\pi}{4} = \sum_{n=0}^{\infty} \arctan\left(\frac{1}{F_{2n+1}}\right).} \tag{6.72}$$

Also, in Problem 3 you will prove the following series for π, due to Castellanos [**47**]:

$$\boxed{\frac{\pi}{\sqrt{5}} = \sum_{n=0}^{\infty} \frac{(-1)^n F_{2n+1} 2^{2n+3}}{(2n+1)(3+\sqrt{5})^{2n+1}}.} \tag{6.73}$$

6.10.3. Machin's arctangent formula for π. In 1706, John Machin (1680–1752) derived a fairly rapid convergent series for π. To derive this expansion, consider the smallest positive angle α whose tangent is 1/5:

$$\tan\alpha = \frac{1}{5} \quad (\text{that is, } \alpha := \arctan(1/5)).$$

Now setting $\theta = \phi = \alpha$ in (6.69), we obtain

$$\tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha} = \frac{2/5}{1-1/25} = \frac{5}{12},$$

so

$$\tan 4\alpha = \frac{2\tan 2\alpha}{1-\tan^2 2\alpha} = \frac{5/6}{1-25/144} = \frac{120}{119},$$

which is just slightly above one. Hence, $4\alpha - \pi/4$ is positive, and moreover,

$$\tan\left(4\alpha - \frac{\pi}{4}\right) = \frac{\tan 4\alpha + \tan\pi/4}{1+\tan 4\alpha\,\tan\pi/4} = \frac{1/119}{1+120/119} = \frac{1}{239}.$$

Taking the inverse tangent of both sides and solving for $\frac{\pi}{4}$, we get

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}.$$

Substituting 1/5 and 1/239 into the Gregory-Madhava series for the inverse tangent, we arrive at Machin's formula for π:

Theorem 6.39 (**Machin's formula**). *We have*

$$\boxed{\pi = 16\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)5^{2n+1}} - 4\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)\,239^{2n+1}}.}$$

Example 6.42. Machin's formula gives many decimal places of π without much effort. Let s_n denote the n-th partial sum of $s := 16\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)5^{2n+1}}$ and t_n that of $t := 4\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)\,239^{2n+1}}$. Then $\pi = s - t$ and by the alternating series error estimate,

$$|s - s_3| \le \frac{16}{9\cdot 5^9} \approx 9.102\times 10^{-7}$$

and

$$|t - t_0| \le \frac{4}{3\cdot(239)^3} \approx 10^{-7}.$$

Therefore,

$$|\pi - (s_3 - t_0)| = |(s - t) - (s_3 - t_0)| \le |s - s_3| + |t - t_0| < 5 \times 10^{-6}.$$

A manageable computation (even without a calculator!) shows that $s_3 - t_0 = 3.14159\ldots$. Therefore, $\pi = 3.14159$ to five decimal places!

Exercises 6.10.

1. From Gregory-Madhava's series, derive the following pretty series

$$\boxed{\frac{\pi}{2\sqrt{3}} = 1 - \frac{1}{3\cdot 3} + \frac{1}{5\cdot 3^2} - \frac{1}{7\cdot 3^3} + \frac{1}{9\cdot 3^4} - + \cdots.}$$

Suggestion: Consider $\arctan(1/\sqrt{3}) = \pi/6$. How many terms of this series do you need to approximate $\pi/2\sqrt{3}$ to within seven decimal places? *History Bite*: Abraham Sharp (1651–1742) used this formula in 1669 to compute π to 72 decimal places, and Thomas Fantet de Lagny (1660–1734) used this formula in 1717 to compute π to 126 decimal places (with a mistake in the 113-th place) [**47**].

2. In this problem we prove (6.72).

(i) Prove that $\arctan\frac{1}{3} = \arctan\frac{1}{5} + \arctan\frac{1}{8}$, and use this prove that

$$\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{5} + \arctan\frac{1}{8}.$$

Prove that $\arctan\frac{1}{8} = \arctan\frac{1}{13} + \arctan\frac{1}{21}$, and use this prove that

$$\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{5} + \arctan\frac{1}{13} + \arctan\frac{1}{21}.$$

From here you can now see the appearance of Fibonacci numbers.

(ii) To continue this by induction, prove that for every natural number n,

$$F_{2n} = \frac{F_{2n+1}F_{2n+2} - 1}{F_{2n+3}}.$$

Suggestion: Can you use (6.71)?

(iii) Using the formula in (ii), prove that

$$\arctan\left(\frac{1}{F_{2n}}\right) = \arctan\left(\frac{1}{F_{2n+1}}\right) + \arctan\left(\frac{1}{F_{2n+2}}\right).$$

Using this formula derive (6.72).

3. In this problem we prove (6.73).

(i) Using (6.70), prove that

$$\tan^{-1}\frac{\sqrt{5}\,x}{1-x^2} = \tan^{-1}\left(\frac{1+\sqrt{5}}{2}\right)x - \tan^{-1}\left(\frac{1-\sqrt{5}}{2}\right)x.$$

(ii) Now prove that

$$\tan^{-1}\frac{\sqrt{5}\,x}{1-x^2} = \sum_{n=0}^{\infty}\frac{(-1)^n F_{2n+1}\, x^{2n+1}}{5^n\,(2n+1)}.$$

(iii) Finally, derive the formula (6.73).

4. In this problem, we prove the breath-taking formula (6.67).

(i) Prove that

$$\frac{\pi}{4} = \sum_{n=1}^{\infty}\Big(\frac{1}{4n-3} - \frac{1}{4n-1}\Big) = \sum_{n=1}^{\infty} f\Big(\frac{1}{n}\Big)$$

where $f(z) = \frac{z}{4-3z} - \frac{z}{4-z}$.

(ii) Use Theorem 6.27 to derive our breath-taking formula.

6.11. ★ Another proof that $\pi^2/6 = \sum_{n=1}^\infty 1/n^2$ (The Basel problem)

Assuming only Gregory-Leibniz-Madhava's series: $\frac{\pi}{4} = \sum_{n=0}^\infty \frac{(-1)^n}{2n+1}$, we give our seventh proof of the fact that[7]

$$\boxed{\frac{\pi^2}{6} = \sum_{n=1}^\infty \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots .}$$

According to Knopp [**120**, p. 324], the proof we are about to give "may be regarded as the most elementary of all known proofs, since it borrows nothing from the theory of functions except the Leibniz series". Knopp attributes the main ideas of the proof to Nicolaus Bernoulli (1687–1759).

6.11.1. Cauchy's arithmetic mean theorem.

Before giving our sixth proof of Euler's sum, we prove the following theorem (attributed to Cauchy by Knopp [**120**, p. 72]).

THEOREM 6.40 (**Cauchy's arithmetic mean theorem**). *If a sequence a_1, a_2, a_3, ... converges to L, then the sequence of arithmetic means (or averages)*

$$m_n := \frac{1}{n}\Big(a_1 + a_2 + \cdots + a_n\Big)$$

also converges to L. Moreover, if the sequence $\{a_n\}$ is nonincreasing, then so is the sequence of arithmetic means $\{m_n\}$.

PROOF. To show that $m_n \to L$, we need to show that

$$m_n - L = \frac{1}{n}\Big((a_1 - L) + (a_2 - L) + \cdots + (a_n - L)\Big)$$

tends to zero as $n \to \infty$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ so that for all $n > N$, we have $|a_n| < \varepsilon/2$. Then for $n > N$, we can write

$$\begin{aligned}|m_n - L| &\le \frac{1}{n}\Big(|(a_1 - L) + \cdots + (a_N - L)|\Big) + \frac{1}{n}\Big(|(a_{N+1} - L) + \cdots + (a_n - L)|\Big)\\ &\le \frac{1}{n}\Big(|(a_1 - L) + \cdots + (a_N - L)|\Big) + \frac{1}{n}\Big(\frac{\varepsilon}{2} + \cdots + \frac{\varepsilon}{2}\Big)\\ &= \frac{1}{n}\Big(|(a_1 - L) + \cdots + (a_N - L)|\Big) + \frac{n-N}{n}\cdot\frac{\varepsilon}{2}\\ &\le \frac{1}{n}\Big(|(a_1 - L) + \cdots + (a_N - L)|\Big) + \frac{\varepsilon}{2}.\end{aligned}$$

By choosing n larger, we can make $\frac{1}{n}\Big(|(a_1 - L) + \cdots + (a_N - L)|\Big)$ also less than $\varepsilon/2$, which shows that $|m_n - L| < \varepsilon$ for n sufficiently large. This shows that $m_n \to L$.

Assume now that $\{a_n\}$ is nonincreasing. We shall prove that $\{m_n\}$ is also nonincreasing; that is, for each n,

$$\frac{1}{n+1}\Big(a_1 + \cdots + a_n + a_{n+1}\Big) \le \frac{1}{n}\Big(a_1 + \cdots + a_n\Big),$$

or, after multiplying both sides by $n(n+1)$,

$$n\Big(a_1 + \cdots + a_n\Big) + na_{n+1} \le n\Big(a_1 + \cdots + a_n\Big) + \Big(a_1 + \cdots + a_n\Big).$$

[7] This proof can be thought of as a systematized version of Problem 3 in Exercises 5.2.

Cancelling, we conclude that the sequence $\{m_n\}$ is nonincreasing if and only if

$$na_{n+1} = \underbrace{a_{n+1} + a_{n+1} + \cdots a_{n+1}}_{n \text{ times}} \leq a_1 + a_2 + \cdots + a_n.$$

But this inequality certainly holds since $a_{n+1} \leq a_k$ for $k = 1, 2, \ldots, n$. This completes the proof. □

There is a related theorem for geometric means found in Problem 2, which can be used to derive the following neat formula:

$$\boxed{e = \lim_{n\to\infty} \left\{ \left(\frac{2}{1}\right)^1 \left(\frac{3}{2}\right)^2 \left(\frac{4}{3}\right)^3 \cdots \left(\frac{n+1}{n}\right)^n \right\}^{1/n}.} \tag{6.74}$$

6.11.2. Proof VII of Euler's formula for $\pi^2/6$. First we shall apply Abel's multiplication theorem to Gregory-Leibniz-Madhava's series:

$$\left(\frac{\pi}{4}\right)^2 = \left(\sum_{n=0}^\infty (-1)^n \frac{1}{2n+1}\right) \cdot \left(\sum_{n=0}^\infty (-1)^n \frac{1}{2n+1}\right).$$

To do so, we first form the n-th term in the Cauchy product:

$$c_n = \sum_{k=0}^n (-1)^k \frac{1}{2k+1} \cdot (-1)^{n-k} \frac{1}{2n-2k+1} = (-1)^n \sum_{k=0}^n \frac{1}{(2k+1)(2n-2k+1)}.$$

Using partial fractions one can check that

$$\frac{1}{(2k+1)(2n-2k+1)} = \frac{1}{2(n+1)} \left(\frac{1}{2k+1} + \frac{1}{2n-2k+1}\right),$$

which implies that

$$c_n = \frac{(-1)^n}{2(n+1)} \left(\sum_{k=0}^n \frac{1}{2k+1} + \sum_{k=0}^n \frac{1}{2n-2k+1}\right) = \frac{(-1)^n}{n+1} \sum_{k=0}^n \frac{1}{2k+1},$$

since $\sum_{k=0}^n \frac{1}{2n-2k+1} = \sum_{k=0}^n \frac{1}{2k+1}$. Thus, we can write

$$\left(\frac{\pi}{4}\right)^2 = \sum_{n=0}^\infty (-1)^n \, m_n \,, \quad \text{where } m_n = \frac{1}{n+1} \left(1 + \frac{1}{3} + \cdots + \frac{1}{2n+1}\right),$$

provided that the series converges! To see that this series converges, note that m_n is exactly the arithmetic mean, or average, of the numbers $1, 1/3, \ldots, 1/(2n+1)$. Since $1/(2n+1) \to 0$ monotonically, Cauchy's arithmetic mean theorem shows that these averages also tend to zero monotonically. In particular, by the alternating series theorem, $\sum_{n=0}^\infty (-1)^n \, m_n$ converges, so by Abel's multiplication theorem, we get (not quite $\pi^2/6$, but pretty nonetheless)

$$\boxed{\frac{\pi^2}{16} = \sum_{n=0}^\infty (-1)^n \frac{1}{n+1} \left(1 + \frac{1}{3} + \cdots + \frac{1}{2n+1}\right).} \tag{6.75}$$

We evaluate the right-hand side using the following theorem (whose proof is technical so you can skip it if you like).

THEOREM 6.41. *Let $\{a_n\}$ be a nonincreasing sequence of positive numbers such that $\sum a_n^2$ converges. Then both series*

$$s := \sum_{n=0}^{\infty}(-1)^n a_n \qquad \text{and} \qquad \delta_k := \sum_{n=0}^{\infty} a_n a_{n+k}, \quad k = 1, 2, 3, \ldots$$

converge. Moreover, $\Delta := \sum_{k=1}^{\infty}(-1)^{k-1}\delta_k$ also converges, and we have the formula

$$\sum_{n=0}^{\infty} a_n^2 = s^2 + 2\Delta.$$

PROOF. Since $\sum a_n^2$ converges, we must have $a_n \to 0$, which implies that $\sum(-1)^n a_n$ converges by the alternating series test. By monotonicity, $a_n a_{n+k} \leq a_n \cdot a_n = a_n^2$ and since $\sum a_n^2$ converges, by comparison, so does each series $\delta_k = \sum_{n=0}^{\infty} a_n a_{n+k}$. Also by monotonicity,

$$\delta_{k+1} = \sum_{n=0}^{\infty} a_n a_{n+k+1} \leq \sum_{n=0}^{\infty} a_n a_{n+k} = \delta_k,$$

so by the alternating series test, the sum Δ converges if $\delta_k \to 0$. To prove that this holds, let $\varepsilon > 0$ and choose N (by invoking the Cauchy criterion for series) such that $a_{N+1}^2 + a_{N+2}^2 + \cdots < \varepsilon/2$. Then, since the sequence $\{a_n\}$ is nondecreasing, we can write

$$\begin{aligned}
\delta_k &= \sum_{n=0}^{\infty} a_n a_{n+k} \\
&= \Big(a_0 a_k + \cdots + a_N a_{N+k}\Big) + \Big(a_{N+1}a_{N+1+k} + a_{N+2}a_{N+2+k} + \cdots\Big) \\
&\leq \Big(a_0 a_k + \cdots + a_N a_k\Big) + \Big(a_{N+1}^2 + a_{N+2}^2 + a_{N+3}^2 + \cdots\Big) \\
&< a_k \cdot \Big(a_0 + \cdots + a_N\Big) + \frac{\varepsilon}{2}.
\end{aligned}$$

As $a_k \to 0$ we can make the first term less than $\varepsilon/2$ for all k large enough. Thus, $\delta_k < \varepsilon$ for all k sufficiently large. This proves that $\delta_k \to 0$ and hence $\Delta = \sum_{k=1}^{\infty}(-1)^{k-1}\delta_k$ converges. Finally, we need to prove the equality

$$\sum_{n=0}^{\infty} a_n^2 = s^2 + 2\Delta = s^2 + 2\sum_{k=1}^{\infty}(-1)^{k-1}\delta_k.$$

To prove this, let s_n denote the n-th partial sum of the series $s = \sum_{n=0}^{\infty}(-1)^n a_n$. We have

$$s_n^2 = \left(\sum_{k=0}^{n}(-1)^k a_k\right)^2 = \sum_{k=0}^{n}\sum_{\ell=0}^{n}(-1)^{k+\ell} a_k\, a_\ell.$$

We can write the double sum on the right as a sum over (k, ℓ) such that $k = \ell$, $k < \ell$, and $\ell < k$:

$$\sum_{k=0}^{n}\sum_{\ell=0}^{n}(-1)^{k+\ell} a_k\, a_\ell = \sum_{k=\ell}(-1)^{k+\ell} a_k\, a_\ell + \sum_{k<\ell}(-1)^{k+\ell} a_k\, a_\ell + \sum_{\ell<k}(-1)^{k+\ell} a_k\, a_\ell,$$

where the smallest k and ℓ can be is 0 and the largest is n. The first sum is just $\sum_{k=0}^n a_k^2$ and by symmetry in k and ℓ, the last two sums are actually the same, so

$$s_n^2 = \sum_{k=0}^n a_k^2 + 2 \sum_{0\le k<\ell\le n} (-1)^{k+\ell} a_k\, a_\ell.$$

In the second sum, $0 \le k < \ell \le n$ so we can write $\ell = k + j$ where $1 \le j \le n$ and $0 \le k \le n - j$. Hence,

$$\sum_{1\le k<\ell\le n} (-1)^{k+\ell} a_k\, a_\ell = \sum_{j=1}^n \sum_{k=0}^{n-j} (-1)^{k+(k+j)} a_k\, a_{k+j} = \sum_{j=1}^n \sum_{k=0}^{n-j} (-1)^j a_k\, a_{k+j}.$$

In summary, we have

$$s_n^2 = \sum_{k=0}^n a_k^2 + 2 \sum_{j=1}^n (-1)^j \left(\sum_{k=0}^{n-j} a_k\, a_{k+j} \right).$$

Let d_n be the n-th partial sum of $2\Delta = 2\sum_{j=1}^\infty (-1)^{j-1}\delta_j$; we need to show that $s_n^2 + d_n \to \sum_{k=0}^\infty a_k^2$ as $n \to \infty$. To this end, we add the expressions for s_n^2 and d_n:

$$\begin{aligned} s_n^2 + d_n &= \sum_{k=0}^n a_k^2 + 2 \sum_{j=1}^n (-1)^j \left(\sum_{k=0}^{n-j} a_k\, a_{k+j} \right) + 2\sum_{j=1}^n (-1)^{j-1}\delta_j \\ &= \sum_{k=0}^n a_k^2 + 2 \sum_{j=1}^n (-1)^j \left(\sum_{k=0}^{n-j} a_k\, a_{k+j} - \delta_j \right). \end{aligned}$$

Recalling that $\delta_j = \sum_{k=0}^\infty a_k a_{k+j}$, we can write $s_n^2 + d_n^2$ as

$$s_n^2 + d_n = \sum_{k=0}^n a_k^2 + 2\sum_{j=1}^n (-1)^j \alpha_j,$$

where

$$\alpha_j := \sum_{k=n-j+1}^\infty a_k a_{k+j} = a_{n-j+1}a_{n+1} + a_{n-j+2}a_{n+2} + a_{n-j+3}a_{n+3} + \cdots .$$

Since the sequence $\{a_n\}$ is nonincreasing, it follows that the sequence $\{\alpha_j\}$ is nondecreasing:

$$\alpha_j = a_{n-j+1}a_{n+1} + a_{n-j+2}a_{n+2} + \cdots \le a_{n-j}a_{n+1} + a_{n-j+1}a_{n+2} + \cdots = \alpha_{j+1}.$$

Now assuming n is even, we have

$$\begin{aligned} \frac{1}{2}\left| s_n^2 + d_n - \sum_{k=0}^n a_k^2 \right| &= \big|(-\alpha_1 + \alpha_2) + (-\alpha_3 + \alpha_4) + \cdots + (-\alpha_{n-1} + \alpha_n)\big| \\ &= (-\alpha_1 + \alpha_2) + (-\alpha_3 + \alpha_4) + \cdots + (-\alpha_{n-1} + \alpha_n) \\ &= -\alpha_1 - (\alpha_3 - \alpha_2) - (\alpha_5 - \alpha_4) - \cdots - (\alpha_{n-1} - \alpha_{n-2}) + \alpha_n \\ &\le \alpha_n = a_1 a_{n+1} + a_2 a_{n+2} + \cdots = \delta_n - a_0 a_n, \end{aligned}$$

where we used the fact that the terms in the parentheses are all nonnegative because the α_j's are nondecreasing. Using a very similar argument, we get

$$\frac{1}{2}\left| s_n^2 + d_n - \sum_{k=0}^n a_k^2 \right| \le \delta_n - a_0 a_n \tag{6.76}$$

for n odd. Therefore, (6.76) holds for all n. We already know that $\delta_n \to 0$ and $a_n \to 0$, so (6.76) shows that the left-hand side tends to zero as $n \to \infty$. This completes the proof of the theorem. □

Finally, we are ready to prove Euler's formula for $\pi^2/6$. To do so, we apply the preceding theorem to the sequence $a_n = 1/(2n+1)$. In this case,

$$\delta_k = \sum_{n=0}^{\infty} a_n a_{n+k} = \sum_{n=0}^{\infty} \frac{1}{(2n+1)(2n+2k+1)}.$$

Writing in partial fractions,

$$\frac{1}{(2n+1)(2n+2k+1)} = \frac{1}{2k}\left\{\frac{1}{2n+1} - \frac{1}{2n+2k+1}\right\},$$

we get (after some cancellations)

$$\delta_k = \frac{1}{2k}\sum_{n=0}^{\infty}\left\{\frac{1}{2n+1} - \frac{1}{2n+2k+1}\right\} = \frac{1}{2k}\left(1 + \frac{1}{3} + \cdots + \frac{1}{2k-1}\right).$$

Hence, the equality $\sum_{n=0}^{\infty} a_n^2 = s^2 + 2\Delta$ takes the form

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \left(\frac{\pi}{4}\right)^2 + \sum_{k=1}^{\infty}(-1)^{k-1}\frac{1}{k}\left(1 + \frac{1}{3} + \cdots \frac{1}{2k-1}\right).$$

However, see (6.75), we already proved that the Cauchy product of Gregory-Leibniz-Madhava's series with itself is given by the sum on the right. Thus,

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \left(\frac{\pi}{4}\right)^2 + \left(\frac{\pi}{4}\right)^2 = \frac{\pi^2}{8}. \tag{6.77}$$

Finally, summing over the even and odd numbers, we have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} + \sum_{n=1}^{\infty} \frac{1}{(2n)^2} = \frac{\pi^2}{8} + \frac{1}{4}\sum_{n=1}^{\infty}\frac{1}{n^2},$$

and solving for $\sum_{n=1}^{\infty} 1/n^2$, we obtain Euler's formula: $\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}$.

EXERCISES 6.11.

1. Find the following limits:

$$(a)\ \lim \frac{1 + 2^{1/2} + 3^{1/3} + \cdots + n^{1/n}}{n},$$

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

2. If a sequence $a_1, a_2, a_3, \ldots$ of positive numbers converges to $L > 0$, prove that the sequence of geometric means $(a_1 a_2 \cdots a_n)^{1/n}$ also converges to L. Suggestion: Take logs of the geometric means. Using this result, prove (6.74). Using (6.74), prove that

$$e = \lim \frac{n}{(n!)^{1/n}}.$$

3. Here is a generalization of Cauchy's arithmetic mean theorem: If a_1, a_2, a_3, $\ldots$ converges to a and b_1, b_2, b_3, $\ldots$ converges to b, then the sequence

$$\frac{1}{n}\Big(a_1 b_n + a_2 b_{n-1} + \cdots + a_{n-1} b_2 + a_n b_1\Big)$$

converges to ab.

CHAPTER 7

More on the infinite: Products and partial fractions

Reason's last step is the recognition that there are an infinite number of things which are beyond it.
Blaise Pascal (1623–1662), Pensees. 1670.

We already met François Viète's infinite product expression for π in Sections 4.10 and 5.1. This chapter is devoted entirely to the theory and application of infinite products, and as a consolation prize we also talk about partial fractions. In Sections 7.1 and 7.2 we present the basics of infinite products. Hold on to your seats, because the rest of the chapter is full of surprises!

We begin with the following "Viète-type" formula for $\log 2$, which is due to Philipp Ludwig von Seidel (1821–1896):

$$\boxed{\log 2 = \frac{2}{1+\sqrt{2}} \cdot \frac{2}{1+\sqrt{\sqrt{2}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{2}}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{\sqrt{2}}}}} \cdots .}$$

In Section 7.3, we give another proof Euler's famous sine formula:

$$\boxed{\sin \pi z = \pi z\Big(1 - \frac{z^2}{1^2}\Big)\Big(1 - \frac{z^2}{2^2}\Big)\Big(1 - \frac{z^2}{3^2}\Big)\Big(1 - \frac{z^2}{4^2}\Big)\Big(1 - \frac{z^2}{5^2}\Big) \cdots ,}$$

In Section 7.4, we look at partial fraction expansions of the trig functions. Recall that if $p(z)$ is a polynomial with roots $r_1, \ldots, r_n$, then we can factor $p(z)$ as $p(z) = a(z - r_1)(z - r_2) \cdots (z - r_n)$, and from elementary calculus, we can write

$$\frac{1}{p(z)} = \frac{1}{a(z - r_1)(z - r_2) \cdots (z - r_n)} = \frac{a_1}{z - r_1} + \frac{a_2}{z - r_2} + \cdots + \frac{a_n}{z - r_n}$$

for some constants $a_1, \ldots, a_n$. You probably studied this in the "partial fraction method of integration" section in your elementary calculus course. Writing

$$\sin \pi z = az(z-1)(z+1)(z-2)(z+2)(z-3)(z+3) \cdots ,$$

Euler thought that we should be able to apply the partial fraction decomposition to $1/\sin \pi z$:

$$\frac{1}{\sin \pi z} = \frac{a_1}{z} + \frac{a_2}{z-1} + \frac{a_3}{z+1} + \frac{a_4}{z-2} + \frac{a_5}{z+2} + \cdots .$$

In Section 7.4, we'll prove that this can be done where $a_1 = 1$ and $a_2 = a_3 = \cdots = -1$. That is, we'll prove that

$$\frac{\pi}{\sin \pi z} = \frac{1}{z} - \frac{1}{z-1} - \frac{1}{z+1} - \frac{1}{z-2} - \frac{1}{z+2} - \frac{1}{z-3} - \frac{1}{z+3} - \cdots .$$

Combining the adjacent factors, $-\frac{1}{z-n} - \frac{1}{z+n} = \frac{2z}{n^2-z^2}$, we get Euler's celebrated partial fraction expansion for sine:

$$\boxed{\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{n^2 - z^2}.} \tag{7.1}$$

We'll also derive partial fraction expansions for the other trig functions. In Section 7.5, we give more proofs of Euler's sum for $\pi^2/6$ using the infinite products and partial fractions we found in Sections 7.3 and 7.4. In Section 7.6, we prove one of the most famous formulas for the Riemann zeta function, namely writing it as an infinite product involving only the *prime* numbers:

$$\boxed{\zeta(z) = \frac{2^z}{2^z-1} \cdot \frac{3^z}{3^z-1} \cdot \frac{5^z}{5^z-1} \cdot \frac{7^z}{7^z-1} \cdot \frac{11^z}{11^z-1} \cdots .}$$

In particular, setting $z = 2$, we get the following expression for $\pi^2/6$:

$$\boxed{\frac{\pi^2}{6} = \prod \frac{p^2}{p^2-1} = \frac{2^2}{2^2-1} \cdot \frac{3^2}{3^2-1} \cdot \frac{5^2}{5^2-1} \cdots .}$$

As a bonus prize, we see how π is related to questions from probability. Finally, in Section 5.3, we derive some awe-inspiring beautiful formulas (too many to list at this moment!). Here are a couple of my favorite formulas of all time:

$$\boxed{\frac{\pi}{4} = \frac{3}{4} \cdot \frac{5}{4} \cdot \frac{7}{8} \cdot \frac{11}{12} \cdot \frac{13}{12} \cdot \frac{17}{16} \cdot \frac{19}{20} \cdot \frac{23}{24} \cdots .}$$

The numerators of the fractions on the right are the odd prime numbers and the denominators are even numbers divisible by four and differing from the numerators by one. The next one is also a beaut:

$$\boxed{\frac{\pi}{2} = \frac{3}{2} \cdot \frac{5}{6} \cdot \frac{7}{6} \cdot \frac{11}{10} \cdot \frac{13}{14} \cdot \frac{17}{18} \cdot \frac{19}{18} \cdot \frac{23}{22} \cdots .}$$

The numerators of the fractions are the odd prime numbers and the denominators are even numbers not divisible by four and differing from the numerators by one.

Chapter 7 objectives: The student will be able to ...

- determine the (absolute) convergence for an infinite product.
- explain the infinite products and partial fraction expansions of the trig functions.
- describe Euler's formulæ for powers of π and their relationship to Riemann's zeta function.

7.1. Introduction to infinite products

We start our journey through infinite products taking careful steps to define what these phenomenal products are.

7.1.1. Basic definitions and examples. Let $\{b_n\}$ be a sequence of complex numbers. An infinite product

$$\prod_{n=1}^{\infty} b_n = b_1 \cdot b_2 \cdot b_3 \cdots$$

is said to **converge** if there exists an $m \in \mathbb{N}$ such that the b_n's are nonzero for all $n \geq m$, and the limit of the partial products $\prod_{k=m}^{n} b_k = b_m \cdot b_{m+1} \cdots b_n$:

$$\lim_{n\to\infty} \prod_{k=m}^{n} b_k = \lim_{n\to\infty} \big(b_m \cdot b_{m+1} \cdots b_n\big) \tag{7.2}$$

converges to a *nonzero* complex value, say p. In this case, we define

$$\prod_{n=1}^{\infty} b_n := b_1 \cdot b_2 \cdots b_{m-1} \cdot p.$$

This definition is of course independent of the m chosen such that the b_n's are nonzero for all $n \geq m$. The infinite product $\prod_{n=1}^{\infty} b_n$ **diverges** if it doesn't converge; that is, either there are infinitely many zero b_n's or the limit (7.2) diverges or the limit (7.2) converges to zero. In this latter case, we say that the infinite product **diverges to zero**. Just as sequences and series can start at any integer, products can also start at any integer: $\prod_{n=k}^{\infty} b_n$, with straightforward modifications of the definition.

Example 7.1. The "harmonic product" $\prod_{n=2}^{\infty}(1-1/n)$ diverges to zero because

$$\prod_{k=2}^{n}\Big(1-\frac{1}{k}\Big) = \Big(1-\frac{1}{2}\Big)\cdots\Big(1-\frac{1}{n}\Big) = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdots\frac{n-1}{n} = \frac{1}{n} \to 0.$$

Example 7.2. On the other hand, the product $\prod_{n=2}^{\infty}(1-1/n^2)$ converges because

$$\begin{aligned}\prod_{k=2}^{n}\Big(1-\frac{1}{k^2}\Big) &= \prod_{k=2}^{n}\frac{k^2-1}{k^2} = \prod_{k=2}^{n}\frac{(k-1)(k+1)}{k\cdot k}\\ &= \frac{1\cdot 3}{2\cdot 2}\cdot\frac{2\cdot 4}{3\cdot 3}\cdot\frac{3\cdot 5}{4\cdot 4}\cdot\frac{4\cdot 6}{5\cdot 5}\cdots\frac{(n-1)(n+1)}{n\cdot n} = \frac{n+1}{2n} \to \frac{1}{2} \neq 0.\end{aligned}$$

Therefore,

$$\prod_{n=2}^{\infty}\Big(1-\frac{1}{n^2}\Big) = \frac{1}{2}.$$

Note that the infinite product $\prod_{n=1}^{\infty}(1-1/n^2)$ also converges, but in this case,

$$\prod_{n=1}^{\infty}\Big(1-\frac{1}{n^2}\Big) := \Big(1-\frac{1}{1^2}\Big)\cdot\lim_{n\to\infty}\prod_{k=2}^{n}\Big(1-\frac{1}{k^2}\Big) = 0\cdot\frac{1}{2} = 0.$$

PROPOSITION 7.1. *If an infinite product converges, then its factors tend to one. Also, a convergent infinite product has the value* 0 *if and only if it has a zero factor.*

PROOF. The second statement is automatic from the definition of convergence. If none of the b_n's vanish for $n \geq m$ and $p_n = b_m \cdot b_{m+1} \cdots b_n$, then $p_n \to p$, a nonzero number, so

$$b_n = \frac{b_m \cdot b_{m+1} \cdots b_{n-1} \cdot b_n}{b_m \cdot b_{m+1} \cdots b_{n-1}} = \frac{p_n}{p_{n-1}} \to \frac{p}{p} = 1.$$

□

Because the factors of a convergent infinite product always tend to one, we henceforth write b_n as $1 + a_n$, so the infinite product takes the form

$$\prod (1 + a_n);$$

then this infinite product converges implies that $a_n \to 0$.

7.1.2. Infinite products and series: the nonnegative case. The following theorem states that the analysis of an infinite product $\prod(1 + a_n)$ with all the a_n's real and nonnegative is completely determined by the infinite series $\sum a_n$.

Theorem 7.2. *An infinite product $\prod(1 + a_n)$ with nonnegative terms a_n converges if and only if the series $\sum a_n$ converges.*

Proof. Let the partial products and partial sums be denoted by

$$p_n = \prod_{k=1}^{n} (1 + a_k) \quad \text{and} \quad s_n = \sum_{k=1}^{n} a_k.$$

Since all the a_k's are nonnegative, both sequences $\{p_n\}$ and $\{s_n\}$ are nondecreasing, so converge if and only if they are bounded. Since $1 \le 1 + x \le e^x$ for any real number x (see Theorem 4.29), it follows that

$$1 \le p_n = \prod_{k=1}^{n} (1 + a_k) \le \prod_{k=1}^{n} e^{a_k} = e^{\sum_{k=1}^{n} a_k} = e^{s_n}.$$

This equation shows that if the sequence $\{s_n\}$ is bounded, then the sequence $\{p_n\}$ is also bounded, and hence converges. Its limit must be ≥ 1, so in particular, is not zero. On the other hand,

$$p_n = (1 + a_1)(1 + a_2) \cdots (1 + a_n) \ge 1 + a_1 + a_2 + \cdots + a_n = 1 + s_n,$$

since the left-hand side, when multiplied out, contains the sum $1 + a_1 + a_2 + \cdots + a_n$ (and a lot of other nonnegative terms too). This shows that if the sequence $\{p_n\}$ is bounded, then the sequence $\{s_n\}$ is also bounded. □

See Problem 4 for the case when the terms a_n are negative.

Example 7.3. Thus, as a consequence of this theorem, the product

$$\prod \left(1 + \frac{1}{n^p}\right)$$

converges for $p > 1$ and diverges for $p \le 1$.

7.1.3. Infinite products for $\log 2$ and e. I found the following gem in [**205**]. Define a sequence $\{e_n\}$ by $e_1 = 1$ and $e_{n+1} = (n+1)(e_n + 1)$ for $n = 1, 2, 3, \ldots$; e.g.

$$e_1 = 1\ ,\ e_2 = 4\ ,\ e_3 = 15\ ,\ e_4 = 64\ ,\ e_5 = 325\ ,\ e_6 = 1956\ , \ldots.$$

Then

$$\boxed{e = \prod_{n=1}^{\infty} \frac{e_n + 1}{e_n} = \frac{2}{1} \cdot \frac{5}{4} \cdot \frac{16}{15} \cdot \frac{65}{64} \cdot \frac{326}{325} \cdot \frac{1957}{1956} \cdots .} \tag{7.3}$$

You will be asked to prove this in Problem 6.

We now prove Philipp Ludwig von Seidel's (1821–1896) formula for $\log 2$:

$$\boxed{\log 2 = \frac{2}{1+\sqrt{2}} \cdot \frac{2}{1+\sqrt{\sqrt{2}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{2}}}} \cdot \frac{2}{1+\sqrt{\sqrt{\sqrt{\sqrt{2}}}}} \cdots .}$$

To prove this, we follow the proof of Viète's formula in Section 5.1.1 using hyperbolic functions instead of trigonometric functions. Let $x \in \mathbb{R}$ be nonzero. Then dividing the identity $\sinh x = 2\cosh(x/2)\sinh(x/2)$ (see Problem 8 in Exercises 4.7) by x, we get

$$\frac{\sinh x}{x} = \cosh(x/2) \cdot \frac{\sinh(x/2)}{x/2}.$$

Replacing x with $x/2$, we get $\sinh(x/2)/(x/2) = \cosh(x/2^2) \cdot \sinh(x/2^2)/(x/2^2)$, therefore

$$\frac{\sinh x}{x} = \cosh(x/2) \cdot \cosh(x/2^2) \cdot \frac{\sinh(x/2^2)}{x/2^2}.$$

Continuing by induction, we obtain for any $n \in \mathbb{N}$,

$$\frac{\sinh x}{x} = \prod_{k=1}^{n} \cosh(x/2^k) \cdot \frac{\sinh(x/2^n)}{x/2^n}.$$

Since $\lim_{z\to 0} \frac{\sinh z}{z} = 1$ (why?), we have $\lim_{n\to\infty} \frac{\sinh(x/2^n)}{x/2^n} = 1$, so taking $n \to \infty$, it follows that

$$\frac{x}{\sinh x} = \lim_{n\to\infty} \prod_{k=1}^{n} \frac{1}{\cosh(x/2^k)}. \tag{7.4}$$

Now let us put $x = \log\theta$, that is, $\theta = e^x$, into the equation (7.4). To this end, observe that

$$\sinh x = \frac{e^x - e^{-x}}{2} = \frac{\theta - \theta^{-1}}{2} = \frac{\theta^2 - 1}{2\theta} \quad\Longrightarrow\quad \frac{x}{\sinh x} = \frac{2\theta\log\theta}{(\theta-1)(\theta+1)}$$

and

$$\cosh(x/2^k) = \frac{e^{\frac{x}{2^k}} + e^{-\frac{x}{2^k}}}{2} = \frac{\theta^{\frac{1}{2^k}} + \theta^{-\frac{1}{2^k}}}{2} = \frac{\theta^{\frac{1}{2^{k-1}}} + 1}{2\theta^{\frac{1}{2^k}}}$$

$$\Longrightarrow\quad \frac{1}{\cosh(x/2^k)} = \frac{2\,\theta^{\frac{1}{2^k}}}{\theta^{1/2^{k-1}} + 1}.$$

Thus,

$$\begin{aligned}
\frac{2\theta\log\theta}{(\theta-1)(\theta+1)} &= \lim_{n\to\infty} \prod_{k=1}^{n} \frac{2\,\theta^{\frac{1}{2^k}}}{\theta^{1/2^{k-1}} + 1} = \lim_{n\to\infty} \left(\prod_{k=1}^{n} \theta^{\frac{1}{2^k}} \cdot \prod_{k=1}^{n} \frac{2}{\theta^{\frac{1}{2^{k-1}}} + 1} \right) \\
&= \lim_{n\to\infty} \left(\theta^{\sum_{k=1}^{n} \frac{1}{2^k}} \cdot \prod_{k=0}^{n-1} \frac{2}{\theta^{\frac{1}{2^k}} + 1} \right).
\end{aligned}$$

Since $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{2^k} = 1$ (this is just the geometric series $\sum_{k=1}^{\infty} \frac{1}{2^k}$), we see that

$$\frac{2\theta\log\theta}{(\theta-1)(\theta+1)} = \theta \cdot \lim_{n\to\infty} \prod_{k=0}^{n-1} \frac{2}{\theta^{\frac{1}{2^k}} + 1} = \theta \cdot \frac{2}{\theta+1} \cdot \lim_{n\to\infty} \prod_{k=1}^{n-1} \frac{2}{\theta^{\frac{1}{2^k}} + 1}.$$

Cancelling like terms, we have, by definition of infinite products, the following beautiful infinite product expansion for $\frac{\log\theta}{\theta-1}$:

$$\boxed{\frac{\log\theta}{\theta-1}=\prod_{k=1}^{\infty}\frac{2}{1+\theta^{\frac{1}{2^k}}}=\frac{2}{1+\sqrt{\theta}}\cdot\frac{2}{1+\sqrt{\sqrt{\theta}}}\cdot\frac{2}{1+\sqrt{\sqrt{\sqrt{\theta}}}}\cdots \textbf{\textit{Seidel's formula}}.}$$

In particular, taking $\theta=2$, we get Seidel's infinite product formula for $\log 2$.

Exercises 7.1.

1. Prove that

$$(a)\ \prod_{n=2}^{\infty}\Big(1+\frac{1}{n^2-1}\Big)=2,\quad (b)\ \prod_{n=3}^{\infty}\Big(1-\frac{2}{n(n-1)}\Big)=\frac{1}{3},$$

$$(c)\ \prod_{n=2}^{\infty}\Big(1+\frac{2}{n^2+n-2}\Big)=3,\quad (d)\ \prod_{n=2}^{\infty}\Big(1+\frac{(-1)^n}{n}\Big)=1.$$

2. Prove that for any $z\in\mathbb{C}$ with $|z|<1$,

$$\prod_{n=0}^{\infty}\Big(1+z^{2^n}\Big)=\frac{1}{1-z}.$$

Conclude that $\prod_{n=0}^{\infty}\Big(1+\Big(\frac{1}{2}\Big)^{2^n}\Big)=2$. Suggestion: Derive, e.g. by induction, a formula for $p_n=\prod_{k=0}^{n}(1+z^{2^k})$ as a geometric sum as in Problem 3e in Exercises 2.2.

3. Determine convergence for:

$$(a)\ \prod_{n=1}^{\infty}\Big(1+\sin^2\Big(\frac{1}{n}\Big)\Big)\quad,\quad (b)\ \prod_{n=1}^{\infty}\Big(1+\Big(\frac{nx^2}{1+n}\Big)^n\Big)\quad,\quad (c)\ \prod_{n=1}^{\infty}\Big(\frac{1+x^2+x^{2n}}{1+x^{2n}}\Big),$$

where for (b) and (c), state for which $x\in\mathbb{R}$, the products converge and diverge.

4. In this problem, we prove that an infinite product $\prod(1-a_n)$ with $0\le a_n<1$ converges if and only if the series $\sum a_n$ converges.
   (i) Let $p_n=\prod_{k=1}^{n}(1-a_k)$ and $s_n=\sum_{k=1}^{n}a_k$. Show that $p_n\le e^{-s_n}$. Conclude that if $\sum a_n$ diverges, then $\prod(1-a_n)$ also diverges (in this case, diverges to zero).
   (ii) Suppose now that $\sum a_n$ converges. Then we can choose m such that $a_m+a_{m+1}+\cdots<1/2$. Prove by induction that

$$(1-a_m)(1-a_{m+1})\cdots(1-a_n)\ge 1-(a_m+a_{m+1}+\cdots+a_n)$$

   for $n=m,m+1,m+2,\ldots$. Conclude that $p_n/p_m\ge 1/2$ for all $n\ge m$, and from this, prove that $\prod(1-a_n)$ converges.
   (iii) For what p is $\prod_{n=2}^{\infty}\big(1-\frac{1}{n^p}\big)$ convergent and divergent?

5. In this problem we derive relationships between series and products. Let $\{a_n\}$ be a sequence of complex numbers with $a_n\neq 0$ for all n.
   (a) Prove that for $n\ge 2$,

$$\prod_{k=1}^{n}(1+a_k)=a_1+\sum_{k=2}^{n}(1+a_1)\cdots(1+a_{k-1})a_k.$$

   Thus, $\prod_{n=1}^{\infty}(1+a_n)$ converges if and only if $a_1+\sum_{k=2}^{\infty}(1+a_1)\cdots(1+a_{k-1})a_k$ converges to a nonzero value, in which case they have the same value.
   (b) Assume that $a_1+\cdots+a_k\neq 0$ for every k. Prove that for $n\ge 2$,

$$\sum_{k=1}^{n}a_k=a_1\prod_{k=2}^{n}\left(1+\frac{a_k}{a_1+a_2+\cdots+a_{k-1}}\right).$$

Thus, $\sum_{n=1}^{\infty} a_n$ converges if and only if $a_1 \prod_{n=2}^{\infty}\left(1+\frac{a_n}{a_1+a_2+\cdots+a_{n-1}}\right)$ either converges or diverges to zero, in which case they have the same value.

(c) Using (b) and the sum $\sum_{n=1}^{\infty} \frac{1}{(n+a-1)(n+a)} = \frac{1}{a}$ from (3.38), prove that

$$\prod_{n=2}^{\infty}\left(1+\frac{a}{(n+a)(n-1)}\right) = a+1.$$

6. In this problem we prove (7.3)
   (i) Let $s_n = \sum_{k=0}^{n} \frac{1}{k!}$. Prove that $e_n = n!\, s_{n-1}$ for $n = 1, 2, \ldots$.
   (ii) Show that $s_n/s_{n-1} = (e_n + 1)/e_n$.
   (iii) Show that $s_n = \prod_{k=1}^{n} \frac{e_k+1}{e_k}$ and then complete the proof. Suggestion: Note that we can write $s_n = (s_1/s_0)\cdot(s_2/s_1)\cdots(s_n/s_{n-1})$.

7.2. Absolute convergence for infinite products

Way back in Section 3.6 we introduced absolute convergence for infinite series and since then we have experienced how incredibly useful this notion is. In this section we continue our study of the basic properties of infinite products by introducing the notion of absolute convergence for infinite products. We also present a general convergence test that is able to test the convergence of any infinite product in terms of a corresponding series of logarithms.

7.2.1. Absolute convergence for infinite products. An infinite product $\prod(1+a_n)$ is said to **converge absolutely** if $\prod(1+|a_n|)$ converges. By Theorem 7.2, $\prod(1+|a_n|)$ converges if and only if $\sum|a_n|$ converges. Therefore, $\prod(1+a_n)$ converges absolutely if and only if the infinite series $\sum a_n$ converges absolutely. We know that if an infinite series is absolutely convergent, then the series itself converges; is this the same for infinite products? The answer is yes, but before proving this we first need the following lemma.

LEMMA 7.3. *Let $\{p_k\}_{k=m}^{\infty}$, where $m \in \mathbb{N}$, be a sequence of complex numbers.*

(a) $\{p_k\}$ converges if and only if the infinite series $\sum_{k=m+1}^{\infty}(p_k - p_{k-1})$ converges, in which case

$$\lim_{k\to\infty} p_k = p_m + \sum_{k=m+1}^{\infty}(p_k - p_{k-1}).$$

(b) If $\{a_j\}_{j=m}^{\infty}$ is a sequence of complex numbers and $p_k = \prod_{j=m}^{k}(1+a_j)$, then

$$|p_k - p_{k-1}| \le |a_k|\, e^{\sum_{j=m}^{k-1}|a_j|}.$$

PROOF. The identity in *(a)* is reminiscent of the telescoping series theorem, Theorem 3.24, and in fact can be derived from it, but let us prove *(a)* directly. To do so, we note that for $k \ge m$, we have

$$p_k = p_m + \sum_{j=m+1}^{k}(p_j - p_{j-1}), \tag{7.5}$$

since the sum on the right telescopes. It follows that the limit $\lim p_k$ exists if and only if the limit $\lim_{k\to\infty}\sum_{j=m+1}^{k}(p_j - p_{j-1})$ exists; in other words, if and only if the infinite series $\sum_{j=m+1}^{\infty}(p_j - p_{j-1})$ converges. In case of convergence, the limit equality in *(a)* follows from taking $k \to \infty$ in (7.5).

To prove *(b)*, observe that

$$\begin{aligned} p_k - p_{k-1} &= \prod_{j=m}^{k}(1+a_j) - \prod_{j=m}^{k-1}(1+a_j) \\ &= (1+a_k)\prod_{j=m}^{k-1}(1+a_j) - \prod_{j=m}^{k-1}(1+a_j) \\ &= a_k \prod_{j=m}^{k-1}(1+a_j). \end{aligned}$$

Therefore, $|p_k - p_{k-1}| \leq |a_k| \prod_{j=m}^{k-1}(1+|a_j|)$. Since $1+x \leq e^x$ for all real numbers x, we have

$$|p_k - p_{k-1}| \leq |a_k| \prod_{j=m}^{k-1} e^{|a_j|} = |a_k|\, e^{\sum_{j=m}^{k-1}|a_j|},$$

just as we wanted. □

Theorem 7.4. *Any absolutely convergent infinite product converges.*

Proof. Let $\prod(1+a_n)$ be absolutely convergent, which is equivalent to the series $\sum|a_n|$ converging; we need to prove that $\prod(1+a_n)$ converges in the usual sense. Since $\sum|a_n|$ converges, by the Cauchy criterion for series we can choose m such that $\sum_{n=m}^{\infty}|a_n| < \frac{1}{2}$. In particular, $|a_k| < 1$ for $k \geq m$, so $1+a_k$ is nonzero for $k \geq m$. For $n \geq m$, let $p_n = \prod_{k=m}^{n}(1+a_k)$. From Lemma 7.3, we know that $\lim p_n$ exists if and only if the infinite series $\sum_{k=m+1}^{\infty}(p_k - p_{k-1})$ converges. To prove that this series converges, note that by *(b)* in Lemma 7.3, for $k > m$ we have

$$|p_k - p_{k-1}| \leq |a_k|\, e^{\sum_{j=m}^{k-1}|a_j|} \quad \Longrightarrow \quad |p_k - p_{k-1}| \leq C\,|a_k|,$$

with $C = e^{1/2}$, recalling that $\sum_{j=m}^{\infty}|a_j| < \frac{1}{2}$. In particular, since $\sum|a_k|$ converges, by the comparison test, the series $\sum_{k=m+1}^{\infty}|p_k - p_{k-1}|$ converges and hence $\sum_{k=m+1}^{\infty}(p_k - p_{k-1})$ also converges. This shows that $\lim p_n$ exists.

We now prove that $\lim p_n \neq 0$. To this end, we claim that for each $n \geq m$, we have

$$|p_n| = \prod_{k=m}^{n}|1+a_k| \geq 1 - \sum_{k=m}^{n}|a_k|. \tag{7.6}$$

We prove (7.6) by induction on $n = m, m+1, m+2, \ldots$. To check the base case, $|1+a_m| \geq 1 - |a_m|$, observe that for any complex number z,

$$1 \leq |1+z-z| \leq |1+z| + |z| \quad \Longrightarrow \quad |1+z| \geq 1 - |z|, \tag{7.7}$$

which in particular proves the base case. Assume that our result holds for $n \geq m$; we prove it for $n+1$. Observe that

$$\begin{aligned}\prod_{k=m}^{n+1} |1+a_k| &= \prod_{k=m}^{n} |1+a_k| \cdot |1+a_{n+1}| \\ &\geq \left(1 - \sum_{k=m}^{n} |a_k|\right)(1-|a_{n+1}|) \quad \text{(induction hypothesis and (7.7))} \\ &= 1 - \sum_{k=m}^{n} |a_k| - |a_{n+1}| + \sum_{k=m}^{n} |a_k||a_{n+1}| \\ &\geq 1 - \sum_{k=m}^{n} |a_k| - |a_{n+1}|,\end{aligned}$$

which is exactly the $n+1$ case. □

Just as for infinite series, the converse of this theorem is not true. For example, the infinite product $\prod_{n=2}^{\infty}\left(1+\frac{(-1)^n}{n}\right)$ converges (and equals 1 — see Problem 1 in Exercises 7.1), but this product is not absolutely convergent.

7.2.2. Infinite products and series: the general case. For nonnegative real numbers $\{a_n\}$, in Theorem 7.2, we showed that the product $\prod(1+a_n)$ converges *if and only if* the series $\sum a_n$ converges. In the general case of a complex sequence $\{a_n\}$, in Theorem 7.4 we showed that the infinite product $\prod(1+a_n)$ still converges *if* the series $\sum |a_n|$ converges. In the general complex case, is there an "if and only if" theorem relating convergence of an infinite product to the convergence of a corresponding infinite series? We now give one such theorem where the series is a series of logarithms. Moreover, we also get a formula for the product in terms of the sum of the infinite series.

THEOREM 7.5. *An infinite product $\prod(1+a_n)$ converges if and only if $a_n \to 0$ and the series*

$$\sum_{n=m+1}^{\infty} \operatorname{Log}(1+a_n),$$

starting from a suitable index $m+1$, converges. Moreover, if L is the sum of the series, then

$$\prod(1+a_n) = (1+a_1)\cdots(1+a_m)\, e^L.$$

PROOF. First of all, we remark that the statement "starting from a suitable index $m+1$" concerning the sum of logarithms is needed because we need to make sure the sum starts sufficiently high so that none of the terms $1+a_n$ is zero (otherwise $\operatorname{Log}(1+a_n)$ is undefined). By Proposition 7.1, in order for the product $\prod(1+a_n)$ to converge, we at least need $a_n \to 0$. Thus, we may assume that $a_n \to 0$; in particular we can fix m such that $n > m$ implies $|a_n| < 1$.

Let $b_n = 1 + a_n$. We shall prove that the infinite product $\prod b_n$ converges if and only if the series

$$\sum_{n=m+1}^{\infty} \operatorname{Log} b_n,$$

converges, and if L is the sum of the series, then

$$\prod b_n = b_1 \cdots b_m \, e^L. \tag{7.8}$$

For $n > m$, let the partial products and partial sums be denoted by

$$p_n = \prod_{k=m+1}^{n} b_k \qquad \text{and} \qquad s_n = \sum_{k=m+1}^{n} \operatorname{Log} b_k.$$

Since $\exp(\operatorname{Log} z) = z$ for any nonzero complex number z, it follows that

$$\exp(s_n) = p_n. \tag{7.9}$$

Thus, if the sum s_n converges to a value L, this equation shows that p_n converges to e^L, which is nonzero, and also proves the formula (7.8).

Conversely, suppose that $\{p_n\}$ converges to a nonzero complex number p. We shall prove that $\{s_n\}$ also converges; once this is established, the formula (7.8) follows from (7.9). Note that replacing b_{m+1} by b_{m+1}/p, we may assume that $p = 1$. For $n > m$, we can write $p_n = \exp(\operatorname{Log} p_n)$, so the formula (7.9) implies that for $n > m$,

$$s_n = \operatorname{Log} p_n + 2\pi i k_n$$

for some integer k_n. Moreover, since

$$s_n - s_{n-1} = \left(\sum_{k=m+1}^{n} \operatorname{Log} b_k \right) - \left(\sum_{k=m+1}^{n-1} \operatorname{Log} b_k \right) = \operatorname{Log} b_n,$$

and $b_n \to 1$ (since $a_n \to 0$), it follows that as $n \to \infty$,

$$\operatorname{Log} p_n - \operatorname{Log} p_{n-1} + 2\pi i(k_n - k_{n-1}) = s_n - s_{n-1} \to \log 1 = 0.$$

By assumption $p_n \to 1$, so we must have $k_n - k_{n-1} \to 0$. Now $k_n - k_{n-1}$ is an integer, so it can approach 0 only if k_n and k_{n-1} are the same integer, say k, for all n sufficiently large. It follows that

$$s_n = \operatorname{Log} p_n + 2\pi i k_n \to \operatorname{Log} 1 + 2\pi i k = 2\pi k,$$

which shows that $\{s_n\}$ converges. □

Exercises 7.2.

1. For what $z \in \mathbb{C}$ are the following products absolutely convergent?

$$(a) \prod_{n=1}^{\infty} \left(1 + z^n\right) \quad , \quad (b) \prod_{n=1}^{\infty} \left(1 + \left(\frac{nz}{1+n}\right)^n\right)$$

$$(c) \prod_{n=1}^{\infty} \left(1 + \sin^2\left(\frac{z}{n}\right)\right) \quad , \quad (d) \prod_{n=2}^{\infty} \left(1 + \frac{z^n}{n \log n}\right) \quad , \quad (e) \prod_{n=1}^{\infty} \frac{\sin(z/n)}{z/n}.$$

2. Here is a nice convergence test: Suppose that $\sum a_n^2$ converges. Then $\prod(1 + a_n)$ converges if and only if the series $\sum a_n$ converges. You may proceed as follows.
   (i) Since $\sum a_n^2$ converges, we know that $a_n \to 0$, so we may henceforth assume that $|a_n|^2 < \frac{1}{2}$ for all n. Prove that

   $$\left| \operatorname{Log}(1 + a_n) - a_n \right| \leq |a_n|^2.$$

   Suggestion: You will need the power series expansion for $\operatorname{Log}(1 + z)$.
   (ii) Prove that $\sum(\operatorname{Log}(1 + a_n) - a_n)$ is absolutely convergent.
   (iii) Prove that $\sum a_n$ converges if and only if $\sum \operatorname{Log}(1 + a_n)$ converges and from this, prove the desired result.

(iv) Does the product $\prod_{n=2}^\infty \big(1 + \frac{(-1)^n}{n}\big)$ converge? What about the product

$$\Big(1+\frac{1}{2}\Big)\Big(1+\frac{1}{3}\Big)\Big(1-\frac{1}{4}\Big)\Big(1+\frac{1}{5}\Big)\Big(1+\frac{1}{6}\Big)\Big(1-\frac{1}{7}\Big)\Big(1+\frac{1}{8}\Big)\Big(1+\frac{1}{9}\Big)\cdots?$$

3. Let $\{a_n\}$ be a sequence of real numbers and assume that $\sum a_n$ converges but $\sum a_n^2$ diverges. In this problem we shall prove that $\prod(1+a_n)$ diverges.
   (i) Prove that there is a constant $C > 0$ such that for all $x \in \mathbb{R}$ with $|x| \le 1/2$, we have
   $$x - \log(1+x) \ge Cx^2.$$
   (ii) Since $\sum a_n$ converges, we know that $a_n \to 0$, so we may assume that $|a_n| \le 1/2$ for all n. Using (i), prove that $\sum \log(1+a_n)$ diverges and hence, $\prod(1+a_n)$ diverges.
   (iii) Does $\prod(1 + \frac{(-1)^{n-1}}{\sqrt{n}})$ converge or diverge?
4. Using the formulas from Problem 5 in Exercises 6.9, prove that for $|z| < 1$,
$$\prod_{n=1}^\infty (1 - z^n) = \exp\left(-\sum_{n=1}^\infty \frac{1}{n}\frac{z^n}{1-z^n}\right) \quad , \quad \prod_{n=1}^\infty (1 + z^n) = \exp\left(\sum_{n=1}^\infty \frac{(-1)^{n-1}}{n}\frac{z^n}{1-z^n}\right).$$
5. In this problem we prove that $\prod(1+a_n)$ is absolutely convergent if and only if the series
$$\sum_{n=m+1}^\infty \mathrm{Log}(1+a_n),$$
starting from a suitable index $m+1$, is absolutely convergent. Proceed as follows.
   (i) Prove that for any complex number z with $|z| \le 1/2$, we have
   $$\frac{1}{2}|z| \le |\,\mathrm{Log}(1+z)| \le \frac{3}{2}|z|. \tag{7.10}$$
   Suggestion: Look at the power series expansion for $\frac{\mathrm{Log}(1+z)}{z} - 1$ and using this power series, prove that for $|z| \le 1/2$, we have $\left|\frac{\mathrm{Log}(1+z)}{z} - 1\right| \le \frac{1}{2}$. Use this inequality to prove (7.10).
   (ii) Now use (7.10) to prove the desired result.

7.3. Euler, Tannery, and Wallis: Product expansions galore

The goal of this section is to learn Tannery's theorem for products and use it to prove Euler's celebrated formula (5.2) stated in the introduction of this chapter:

THEOREM 7.6 (**Euler's product for sine**). *For any complex z, we have*
$$\boxed{\sin \pi z = \pi z \prod_{n=1}^\infty \Big(1 - \frac{z^2}{n^2}\Big).}$$

We give two proofs of this astounding result. We also prove Wallis' infinite product expansion for π. To begin, we first need

7.3.1. Tannery's theorem for products.

THEOREM 7.7 (**Tannery's theorem for infinite products**). *For each natural number n, let $\prod_{k=1}^{m_n}(1+a_k(n))$ be a finite product where $m_n \to \infty$ as $n \to \infty$. If for each k, $\lim_{n\to\infty} a_k(n)$ exists, and there is a series $\sum_{k=1}^\infty M_k$ of nonnegative real numbers such that $|a_k(n)| \le M_k$ for all k, n, then*
$$\lim_{n\to\infty} \prod_{k=1}^{m_n} (1 + a_k(n)) = \prod_{k=1}^\infty \lim_{n\to\infty} (1 + a_k(n));$$

that is, both sides are well-defined (the limits and products converge) and are equal.

Proof. First of all, we remark that the infinite product on the right converges. Indeed, if we put $a_k := \lim_{n\to\infty} a_k(n)$, which exists by assumption, then taking $n \to \infty$ in the inequality $|a_k(n)| \le M_k$, we have $|a_k| \le M_k$ as well. Therefore, by the comparison test, the series $\sum_{k=1}^\infty a_k$ converges absolutely and hence, by Theorem 7.4, the infinite product $\prod_{k=1}^\infty (1+a_k)$ converges.

Now to our proof. Since $\sum M_k$ converges, $M_k \to 0$, so we can choose $m \in \mathbb{N}$ such that for all $k \ge m$, we have $M_k < 1$. This implies that $|a_k| < 1$ for $k \ge m$, so $1 + a_k$ is nonzero for $k \ge m$. Put

$$p(n) = \prod_{k=1}^{m_n} (1 + a_k(n))$$

and, for n large enough so that $m_n > m$, write

$$p(n) = q(n) \cdot \prod_{k=m}^{m_n} (1 + a_k(n)), \text{ where } \quad q(n) = \prod_{k=1}^{m-1} (1 + a_k(n)).$$

Since $q(n)$ is a finite product, $q(n) \to \prod_{k=1}^{m-1}(1+a_k)$ as $n \to \infty$; therefore we just have to prove that

$$\lim_{n\to\infty} \prod_{k=m}^{m_n} (1 + a_k(n)) = \prod_{k=m}^{\infty} (1 + a_k).$$

Consider the partial products

$$p_k(n) = \prod_{j=m}^{k} (1 + a_j(n)) \quad \text{and} \quad p_k = \prod_{j=m}^{k} (1 + a_j).$$

Since these are finite products and $a_j = \lim_{n\to\infty} a_j(n)$, by the algebra of limits we have $\lim_{n\to\infty} p_k(n) = p_k$. Now observe that

$$\prod_{j=m}^{m_n} (1 + a_j(n)) = p_{m_n}(n) = p_m(n) + \sum_{k=m}^{m_n} (p_k(n) - p_{k-1}(n)),$$

since the right-hand side telescopes to $p_{m_n}(n)$, and by the limit identity in *(a)* of Lemma 7.3, we know that

$$\prod_{j=m}^{\infty} (1 + a_j) = p_m + \sum_{k=m+1}^{\infty} (p_k - p_{k-1}),$$

since $\prod_{j=m}^\infty (1+a_j) := \lim_{k\to\infty} p_k$. Also, by Part *(b)* of Lemma 7.3, we have

$$|p_k(n) - p_{k-1}(n)| \le |a_k(n)|\, e^{\sum_{j=m}^{k-1} |a_j(n)|} \le M_k\, e^{\sum_{j=m}^{k-1} M_j} \le C M_k,$$

where $C = e^{\sum_{j=m}^\infty M_j}$. Since $\sum_{k=m+1}^\infty C M_k$ converges, by Tannery's theorem for series we have

$$\begin{aligned}\lim_{n\to\infty} \sum_{k=m+1}^{m_n} (p_k(n) - p_{k-1}(n)) &= \sum_{k=m+1}^{\infty} \lim_{n\to\infty} (p_k(n) - p_{k-1}(n)) \\ &= \sum_{k=m+1}^{\infty} (p_k - p_{k-1}).\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\prod_{j=m}^{m_n}(1+a_j(n)) &= \lim_{n\to\infty}\left(p_m(n)+\sum_{k=m+1}^{m_n}(p_k(n)-p_{k-1}(n))\right)\\ &= p_m+\lim_{n\to\infty}\sum_{k=m+1}^{m_n}(p_k(n)-p_{k-1}(n))\\ &= p_m+\sum_{k=m+1}^{\infty}(p_k-p_{k-1})=\prod_{j=m}^{\infty}(1+a_j),\end{aligned}$$

which is what we wanted to prove. □

See Problem 6 for another (shorter) proof using complex logarithms.

7.3.2. Expansion of sine III. (Cf. [**41**, p. 294]). Our third proof of Euler's infinite product for sine is a Tannery's theorem version of the proof found in Section 5.1 of Chapter 5. To this end, first recall from Lemma 5.1 of that section and the work done in that section, that for any $z\in\mathbb{C}$, we have

$$\sin z=\lim_{n\to\infty}F_n(z),$$

where $n=2m+1$ is odd and

$$F_n(z)=z\prod_{k=1}^{m}\left(1-\frac{z^2}{n^2\tan^2(k\pi/n)}\right).$$

Thus,

$$\begin{aligned}\sin z &= \lim_{m\to\infty}\left\{z\prod_{k=1}^{m}\left(1-\frac{z^2}{n^2\tan^2(k\pi/n)}\right)\right\}\\ &= \lim_{m\to\infty}z\prod_{k=1}^{m}(1+a_k(m)),\end{aligned}$$

where $a_k(m):=-\frac{z^2}{n^2\tan^2(k\pi/n)}$ with $n=2m+1$. Second, since $\lim_{z\to0}\frac{\tan z}{z}=\lim_{z\to0}\frac{\sin z}{z}\cdot\frac{1}{\cos z}=1$, we see that

$$\begin{aligned}\lim_{m\to\infty}a_k(m) &= \lim_{m\to\infty}-\frac{z^2}{(2m+1)^2\tan^2(k\pi/(2m+1))}\\ &= \lim_{m\to\infty}-\frac{z^2}{k^2\pi^2\left(\frac{\tan(k\pi/(2m+1))}{k\pi/(2m+1)}\right)^2}=-\frac{z^2}{k^2\pi^2}.\end{aligned}$$

Third, in Lemma 4.55 back in Section 4.10, we proved that

$$x<\tan x,\qquad \text{for } 0<x<\pi/2. \tag{7.11}$$

In particular, for any $z\in\mathbb{C}$, if $n=2m+1$ and $1\le k\le m$, then

$$\left|\frac{z^2}{n^2\tan^2(k\pi/n)}\right|\le\frac{|z|^2}{n^2(k\pi)^2/n^2}=\frac{|z|^2}{k^2\pi^2}=:M_k.$$

Thus, for all k, m, $|a_k(m)| \le M_k$. Finally, since the sum $\sum_{k=1}^\infty M_k$ converges, by Tannery's theorem for infinite products, we have

$$\sin z = \lim_{m\to\infty} z \prod_{k=1}^{m} (1 + a_k(m)) = z \prod_{k=1}^{\infty} \lim_{m\to\infty} (1 + a_k(m)) = z \prod_{k=1}^{\infty} \left(1 - \frac{z^2}{k^2\pi^2}\right).$$

After replacing z by πz, we get Euler's infinite product expansion for $\sin \pi z$. This completes Proof III of Theorem 7.6. In particular, we see that

$$\pi i \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2}\right) = \pi i \prod_{k=1}^{\infty} \left(1 - \frac{i^2}{k^2}\right) = \sin \pi i = \frac{e^{-\pi} - e^{\pi}}{2i}.$$

Thus, we have derived the very pretty formula

$$\boxed{\frac{e^{\pi} - e^{-\pi}}{2\pi} = \prod_{n=1}^{\infty} \left(1 + \frac{1}{n^2}\right).}$$

Recall from Section 7.1 how easy it was to find that $\prod_{n=1}^\infty \left(1 - \frac{1}{n^2}\right) = 1/2$, but replacing $-1/n^2$ with $+1/n^2$ is a whole different story!

7.3.3. Expansion of sine IV. Our fourth proof of Euler's infinite product for sine is based on the following neat identity involving sines instead of tangents!

Lemma 7.8. *If $n = 2m + 1$ with $m \in \mathbb{N}$, then for any $z \in \mathbb{C}$,*

$$\sin nz = n \sin z \prod_{k=1}^{m} \left(1 - \frac{\sin^2 z}{\sin^2(k\pi/n)}\right).$$

Proof. Lemma 2.26 shows that for each $k \in \mathbb{N}$, $2\cos kz$ is a polynomial in $2\cos z$ of degree k (with integer coefficients, although this fact is not important for this lemma). Technically speaking, Lemma 2.26 was proved under the assumption that z is real, but the proof only used the angle addition formula for cosine, which holds for complex variables as well. Any case, since $2\cos kz$ is a polynomial in $2\cos z$ of degree k, it follows that $\cos kz$ is a polynomial in $\cos z$ of degree k, say $\cos kz = Q_k(\cos z)$ where Q_k is a polynomial of degree k. In particular,

$$\cos 2kz = Q_k(\cos 2z) = Q_k(1 - 2\sin^2 z),$$

so $\cos 2kz$ is a polynomial of degree k in $\sin^2 z$. Now using the addition formulas for sine, we get, for each $k \in \mathbb{N}$,

$$\sin(2k+1)z - \sin(2k-1)z = 2\sin z \cdot \cos(2kz) = 2\sin z \cdot Q_k(1 - 2\sin^2 z). \tag{7.12}$$

We claim that for any $m = 0, 1, 2, \ldots$, we have

$$\sin(2m+1)z = \sin z \cdot P_m(\sin^2 z), \tag{7.13}$$

where P_m is a polynomial of degree m. For example, if $m = 0$, then $\sin z = \sin z \cdot P_0(\sin^2 z)$ where $P_0(w) = 1$ is the constant polynomial 1. If $m = 1$, then by (7.12) with $k = 1$, we have

$$\begin{aligned} \sin(3z) &= \sin z + 2\sin z \cdot Q_1(1 - 2\sin^2 z) \\ &= \sin z\big(1 + 2\sin z \cdot Q_1(1 - 2\sin^2 z)\big) = \sin z \cdot P_1(\sin^2 z), \end{aligned}$$

where $P_1(w) = 1 + 2Q_1(1 - 2w)$. To prove (7.13) for general m just requires an induction argument based on (7.12), which we leave to the interested reader. Now, observe that $\sin(2m+1)z$ is zero when $z = z_k$ with $z_k = k\pi/(2m+1)$ where

$k = 1, 2, \ldots, m$. Also observe that since $0 < z_1 < z_2 < \cdots < z_m < \pi/2$, the m values $\sin z_k$ are distinct positive values. Hence, according to (7.13), $P_m(w) = 0$ at the m distinct values $w = \sin^2 z_k$, $k = 1, 2, \ldots, m$. Thus, as a consequence of the fundamental theorem of algebra, the polynomial $P_m(w)$ can be factored into a constant times

$$(w - z_1)(w - z_2) \cdots (w - z_m) = \prod_{k=1}^{m} \left(w - \sin^2 \left(\frac{k\pi}{2m+1} \right) \right) = \prod_{k=1}^{m} \left(w - \sin^2 \left(\frac{k\pi}{n} \right) \right),$$

(since $n = 2m + 1$) which is a constant times

$$\prod_{k=1}^{m} \left(1 - \frac{w}{\sin^2(k\pi/n)} \right).$$

Setting $w = \sin^2 z$, we obtain

$$\sin(2m+1)z = \sin z \cdot P_m(\sin^2 z) = a \sin z \cdot \prod_{k=1}^{m} \left(1 - \frac{\sin^2 z}{\sin^2(k\pi/n)} \right),$$

for some constant a. Since $\sin(2m + 1)z/ \sin z$ has limit equal to $2m + 1$ as $z \to 0$, it follows that $a = 2m + 1$. This completes the proof of the lemma. □

We are now ready to give our fourth proof of Euler's infinite product for sine. To this end, we let $n \geq 3$ be odd and we replace z by z/n in Lemma 7.8 to get

$$\sin z = n \sin(z/n) \prod_{k=1}^{m} \left(1 - \frac{\sin^2(z/n)}{\sin^2(k\pi/n)} \right),$$

where $n = 2m + 1$. Since

$$\lim_{m\to\infty} (2m + 1) \sin(z/(2m + 1)) = \lim_{m\to\infty} z \frac{\sin(z/(2m + 1))}{z/(2m + 1)} = z,$$

we have

$$\sin z = z \lim_{m\to\infty} \prod_{k=1}^{m} \left(1 - \frac{\sin^2(z/n)}{\sin^2(k\pi/n)} \right) = z \lim_{m\to\infty} \prod_{k=1}^{m} (1 + a_k(m))$$

where $a_k(m) := -\frac{\sin^2(z/n)}{\sin^2(k\pi/n)}$ with $n = 2m+1$. Since we are taking $m \to \infty$, we can always make sure that $n = 2m + 1 > |z|$, which we henceforth assume. Now recall from Lemmas 5.6 and 5.7 that there is a constant $c > 0$ such that for any $0 \leq x \leq \frac{\pi}{2}$, we have $c\,x \leq \sin x$, and for any $w \in \mathbb{C}$ with $|w| \leq 1$, we have $|\sin w| \leq \frac{6}{5}|w|$. It follows that for any $k = 1, 2, \ldots, m$,

$$\left| \frac{\sin^2(z/n)}{\sin^2(k\pi/n)} \right| \leq \frac{(6/5|z/n|)^2}{c^2(k\pi/n)^2} = \frac{36|z|^2}{25c^2\pi^2} \cdot \frac{1}{k^2} =: M_k,$$

Since the sum $\sum_{k=1}^\infty M_k$ converges, and

$$\begin{aligned}\lim_{m\to\infty} a_k(m) &= -\lim_{m\to\infty} \frac{\sin^2(z/(2m+1))}{\sin^2(k\pi/(2m+1))}\\ &= -\lim_{m\to\infty} \frac{z^2}{k^2\pi^2}\cdot\frac{\left(\frac{\sin(z/(2m+1))}{z/(2m+1)}\right)^2}{\left(\frac{\sin(k\pi/(2m+1))}{k\pi/(2m+1)}\right)^2} = -\frac{z^2}{k^2\pi^2},\end{aligned}$$

Tannery's theorem for infinite products implies that

$$\sin z = z\lim_{m\to\infty}\prod_{k=1}^m (1+a_k(m)) = z\prod_{k=1}^\infty \lim_{m\to\infty}(1+a_k(m)) = z\prod_{k=1}^\infty\left(1-\frac{z^2}{k^2\pi^2}\right).$$

Finally, replacing z by πz completes Proof IV of Euler's product formula.

7.3.4. Euler's cosine expansion. We can derive an infinite product expansion for the cosine function easily from the sine expansion. In fact, using the double angle formula for sine, we get

$$\cos\pi z = \frac{\sin 2\pi z}{2\sin\pi z} = \frac{2\pi z\cdot\prod_{n=1}^\infty\left(1-\frac{4z^2}{n^2}\right)}{2\pi z\cdot\prod_{n=1}^\infty\left(1-\frac{z^2}{n^2}\right)} = \frac{\prod_{n=1}^\infty\left(1-\frac{4z^2}{n^2}\right)}{\prod_{n=1}^\infty\left(1-\frac{z^2}{n^2}\right)}.$$

The top product can be split as a product of even and odd terms:

$$\prod_{n=1}^\infty\left(1-\frac{4z^2}{(2n-1)^2}\right)\prod_{n=1}^\infty\left(1-\frac{4z^2}{(2n)^2}\right) = \prod_{n=1}^\infty\left(1-\frac{4z^2}{(2n-1)^2}\right)\prod_{n=1}^\infty\left(1-\frac{z^2}{n^2}\right),$$

from which we get (see Problem 3 for three more proofs)

$$\boxed{\cos\pi z = \prod_{n=1}^\infty\left(1-\frac{4z^2}{(2n-1)^2}\right).}$$

Exercises 7.3.

1. Put $z=\pi/4$ into the cosine expansion to derive the following elegant product for $\sqrt{2}$:

$$\boxed{\sqrt{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{10}{9}\cdot\frac{10}{11}\cdots.}$$

Compare this with Wallis' formula:

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdot\frac{8}{9}\cdot\frac{10}{9}\cdot\frac{10}{11}\cdots.$$

Thus, the product for $\sqrt{2}$ is obtained from Wallis' formula for $\pi/2$ by removing the factors with numerators that are multiples of 4.

2. Prove that

$$\sinh\pi z = \pi z\prod_{k=1}^\infty\left(1+\frac{z^2}{k^2}\right) \quad\text{and}\quad \cosh\pi z = \prod_{n=1}^\infty\left(1+\frac{z^2}{(2n-1)^2}\right).$$

3. (**Euler's infinite product for** $\cos\pi z$) Here are three more proofs!

(a) Replace z by $-z+1/2$ in the sine product to derive the cosine product. Suggestion: Begin by showing that

$$\left(1-\frac{(-z+\frac{1}{2})^2}{n^2}\right)=\left(1-\frac{1}{4n^2}\right)\cdot\left(1+\frac{2z}{2n-1}\right)\left(1-\frac{2z}{2n+1}\right).$$

(b) For our second proof, show that for n even, we can write

$$\cos z=\prod_{k=1}^{n-1}\left(1-\frac{\sin^2(z/n)}{\sin^2(k\pi/2n)}\right),\quad k=1,3,5,\dots,n-1.$$

Using Tannery's theorem, deduce the cosine expansion.

(c) Write $\cos z=\lim_{n\to\infty}G_n(z)$, where

$$G_n(z)=\frac{1}{2}\left\{\left(1+\frac{iz}{n}\right)^n+\left(1-\frac{iz}{n}\right)^n\right\}.$$

Prove that if $n=2m$ with $m\in\mathbb{N}$, then

$$G_n(z)=\prod_{k=0}^{m}\left(1-\frac{z^2}{n^2\tan^2((2k+1)\pi/(2n))}\right).$$

Using Tannery's theorem, deduce the cosine expansion.

4. Prove that

$$1-\sin z=\left(1-\frac{2z}{1}\right)^2\left(1+\frac{2z}{3}\right)^2\left(1-\frac{2z}{5}\right)^2\left(1+\frac{2z}{7}\right)^2\cdots$$

Suggestion: First show that $1-\sin z=2\sin^2(\frac{\pi}{4}-\frac{z}{2})$.

5. Determine the following limits.

$$(a)\ \lim_{n\to\infty}\left\{\left(1-\frac{1}{4n^2\log\left(1+\left(\frac{2}{2n}\right)^2\right)}\right)\cdot\left(1-\frac{1}{4n^2\log\left(1+\left(\frac{3}{2n}\right)^2\right)}\right)\cdot\right.$$
$$\left.\left(1-\frac{1}{4n^2\log\left(1+\left(\frac{4}{2n}\right)^2\right)}\right)\cdots\left(1-\frac{1}{4n^2\log\left(1+\left(\frac{n}{2n}\right)^2\right)}\right)\right\},$$

$$(b)\ \lim_{n\to\infty}\left\{\left(1+\frac{1}{4n^2\sin\left(\frac{4\cdot1^2-1}{4n^2-1}\right)}\right)\cdot\left(1+\frac{1}{4n^2\sin\left(\frac{4\cdot2^2-1}{4n^2-1}\right)}\right)\cdots\left(1+\frac{1}{4n^2\sin\left(\frac{4\cdot n^2-1}{4n^2-1}\right)}\right)\right\}.$$

6. In this problem we prove Tannery's theorem for products using complex logarithms. Assume the hypotheses and notations of Theorem 7.7. Since $\sum M_k$ converges, $M_k\to 0$, so we can choose m such that for all $k\geq m$, we have $M_k<1/2$. Then as in the proof of Theorem 7.7, we just have to show that

$$\lim_{n\to\infty}\prod_{k=m}^{m_n}(1+a_k(n))=\prod_{k=m}^{\infty}(1+a_k). \tag{7.14}$$

(i) Show that Tannery's theorem for series implies that

$$\lim_{n\to\infty}\sum_{k=m}^{m_n}\operatorname{Log}(1+a_k(n))=\sum_{k=m}^{\infty}\operatorname{Log}(1+a_k).$$

Suggestion: Use the inequality (7.10) in Problem 5 of Exercises 7.2.

(ii) From (i), deduce (7.14).

7. (**Tannery's theorem II**) For each natural number n, let $\prod_{k=1}^{\infty}(1+a_k(n))$ be a convergent infinite product. If for each k, $\lim_{n\to\infty} a_k(n)$ exists, and there is a series $\sum_{k=1}^{\infty} M_k$ of nonnegative real numbers such that $|a_k(n)| \le M_k$ for all k, n, prove that

$$\lim_{n\to\infty} \prod_{k=1}^{\infty}(1+a_k(n)) = \prod_{k=1}^{\infty} \lim_{n\to\infty}(1+a_k(n));$$

that is, both sides are well-defined (the limits and products converge) and are equal.

7.4. Partial fraction expansions of the trigonometric functions

The goal of this section is to prove Euler's partial fraction expansion (7.1):

Theorem 7.9 (**Euler's partial fraction** $(\frac{\pi}{\sin \pi z})$). *We have*

$$\boxed{\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{n^2 - z^2} \quad \textit{for all } z \in \mathbb{C} \setminus \mathbb{Z}.}$$

We also derive partial fraction expansions for the other trigonometric functions. We begin with the cotangent.

7.4.1. Partial fraction expansion of the cotangent. We shall prove the following theorem (from which we'll derive the sine expansion).

Theorem 7.10 (**Euler's partial fraction** $(\pi z \cot \pi z)$). *We have*

$$\boxed{\pi z \cot \pi z = 1 + 2z^2 \sum_{n=1}^{\infty} \frac{1}{z^2 - n^2} \quad \textit{for all } z \in \mathbb{C} \setminus \mathbb{Z}.}$$

Our proof of Euler's expansion of the cotangent is based on the following lemma.

Lemma 7.11. *For any noninteger complex number z and $n \in \mathbb{N}$, we have*

$$\pi z \cot \pi z = \frac{\pi z}{2^n} \cot \frac{\pi z}{2^n} + \sum_{k=1}^{2^{n-1}-1} \frac{\pi z}{2^n}\left(\cot \frac{\pi(z+k)}{2^n} + \cot \frac{\pi(z-k)}{2^n}\right) - \frac{\pi z}{2^n} \tan \frac{\pi z}{2^n}.$$

Proof. Using the double angle formula

$$2 \cot 2z = 2 \frac{\cos 2z}{\sin 2z} = \frac{\cos^2 z - \sin^2 z}{\cos z \sin z} = \cot z - \tan z,$$

we see that

$$\cot 2z = \frac{1}{2}\Big(\cot z - \tan z\Big).$$

Replacing z with $\pi z/2$, we get

$$\cot \pi z = \frac{1}{2}\left(\cot \frac{\pi z}{2} - \tan \frac{\pi z}{2}\right). \tag{7.15}$$

Multiplying this equality by πz proves our lemma for $n = 1$. In order to proceed by induction, we note that since $\tan z = -\cot(z \pm \pi/2)$, we find that

$$\cot \pi z = \frac{1}{2}\left(\cot \frac{\pi z}{2} + \cot \frac{\pi(z \pm 1)}{2}\right). \tag{7.16}$$

This is the main formula on which induction may be applied to prove our lemma. For instance, let's take the case $n = 2$. Considering the positive sign in the second cotangent, we have

$$\cot \pi z = \frac{1}{2}\bigg(\cot \frac{\pi z}{2} + \cot \frac{\pi(z+1)}{2} \bigg).$$

Applying (7.16) to each cotangent on the right of this equation, using the plus sign for the first and the minus sign for the second, we get

$$\begin{aligned} \cot \pi z &= \frac{1}{2^2}\left\{ \left(\cot \frac{\pi z}{2^2} + \cot \frac{\pi(\frac{z}{2}+1)}{2} \right) + \left(\cot \frac{\pi(z+1)}{2^2} + \cot \frac{\pi(\frac{z+1}{2}-1)}{2} \right) \right\} \\ &= \frac{1}{2^2}\left\{ \cot \frac{\pi z}{2^2} + \cot \frac{\pi(z+2)}{2^2} + \cot \frac{\pi(z+1)}{2^2} + \cot \frac{\pi(z-1)}{2^2} \right\}, \end{aligned}$$

which, after bringing the second cotangent to the end, takes the form

$$\cot \pi z = \frac{1}{2^2}\left\{ \cot \frac{\pi z}{2^2} + \cot \frac{\pi(z+1)}{2^2} + \cot \frac{\pi(z-1)}{2^2} + \cot \left(\frac{\pi z}{2^2} + \frac{\pi}{2} \right) \right\}.$$

However, the last term is exactly $-\tan \pi z/2^2$, and so our lemma is proved for $n = 2$. Continuing by induction proves our lemma for general n. □

Fix a noninteger z; we shall prove Euler's expansion for the cotangent. Note that $\lim_{n\to\infty} \frac{\pi z}{2^n} \tan(\frac{\pi z}{2^n}) = 0 \cdot \tan 0 = 0$, and since

$$\lim_{w\to 0} w \cot w = \lim_{w\to 0} \frac{w}{\sin w} \cdot \cos w = 1 \cdot 1 = 1, \tag{7.17}$$

we have $\lim_{n\to\infty} \frac{\pi z}{2^n} \cot \frac{\pi z}{2^n} = 1$. Therefore, taking $n \to \infty$ in the formula from the preceding Lemma 7.11, we conclude that

$$\begin{aligned} \pi z \cot \pi z &= 1 + \lim_{n\to\infty} \left\{ \sum_{k=1}^{2^{n-1}-1} \frac{\pi z}{2^n} \left(\cot \frac{\pi(z+k)}{2^n} + \cot \frac{\pi(z-k)}{2^n} \right) \right\} \\ &= 1 + \lim_{n\to\infty} \sum_{k=1}^{2^{n-1}-1} a_k(n), \end{aligned}$$

where

$$a_k(n) = \frac{\pi z}{2^n} \left(\cot \frac{\pi(z+k)}{2^n} + \cot \frac{\pi(z-k)}{2^n} \right).$$

We shall apply Tannery's theorem to this sum. To this end, observe that, from (7.17),

$$\lim_{n\to\infty} \frac{\pi z}{2^n} \cot \frac{\pi(z+k)}{2^n} = \frac{z}{z+k} \lim_{n\to\infty} \frac{\pi(z+k)}{2^n} \cot \frac{\pi(z+k)}{2^n} = \frac{z}{z+k},$$

and in a similar way,

$$\lim_{n\to\infty} \frac{\pi z}{2^n} \cot \frac{\pi(z-k)}{2^n} = \frac{z}{z-k}.$$

Thus,

$$\lim_{n\to\infty} a_k(n) = \frac{z}{z+k} + \frac{z}{z-k} = \frac{2z^2}{z^2-k^2},$$

so Tannery's theorem gives Euler's cotangent expansion:

$$\pi z \cot \pi z = 1 + 2z^2 \sum_{k=1}^{\infty} \frac{1}{z^2-k^2},$$

provided of course we can show that $|a_k(n)| \leq M_k$ where $\sum M_k < \infty$. Actually, we shall prove that are $m, N \in \mathbb{N}$ such that $|a_k(n)| \leq M_k$ for all $n > N$ and $k \geq m$ where $\sum_{k=m}^{\infty} M_k < \infty$. The conclusion of Tannery's theorem will still hold with these conditions. (Why so?)

To bound each $a_k(n)$, we use the formula

$$\cot(\alpha + \beta) + \cot(\alpha - \beta) = \frac{\sin 2\alpha}{\sin^2 \alpha - \sin^2 \beta}.$$

This formula is obtained by expressing $\cot(\alpha \pm \beta)$ in terms of cosine and sine and using the angle addition formulas (the diligent reader will supply the details!). Setting $\alpha = \pi z/2^n$ and $\beta = \pi k/2^n$, we obtain

$$\cot \frac{\pi(z+k)}{2^n} + \cot \frac{\pi(z-k)}{2^n} = \frac{\sin 2\alpha}{\sin^2 \alpha - \sin^2 \beta}, \tag{7.18}$$

where we keep the notation $\alpha = \pi z/2^n$ and $\beta = \pi k/2^n$ on the right. Our goal now is to bound the term on the right of (7.18). Choose $N \in \mathbb{N}$ such that for all $n > N$, we have $|\alpha| = |\pi z/2^n| < 1/2$. Then for $n > N$, according to Lemma 5.7,

$$|\sin 2\alpha| \leq \frac{6}{5}|2\alpha| \leq 3|\alpha| \quad \text{and} \quad |\sin \alpha| \leq \frac{6}{5}|\alpha| \leq 2|\alpha|,$$

and, since $\beta = \pi k/2^n < \pi/2$ for $k = 1, \ldots, 2^{n-1} - 1$, according to Lemma 5.6, for some $c > 0$,

$$c\,\beta \leq \sin \beta.$$

Hence, for $n > N$,

$$c^2\beta^2 \leq \sin^2 \beta \leq |\sin^2 \alpha - \sin^2 \beta| + |\sin^2 \alpha| \leq |\sin^2 \alpha - \sin^2 \beta| + 4|\alpha|^2$$
$$\Longrightarrow \quad c^2\beta^2 - 4|\alpha|^2 \leq |\sin^2 \alpha - \sin^2 \beta|.$$

Choose $m \in \mathbb{N}$ such that $c\,m > 2|z|$. Then for $k \geq m$, we have

$$c^2\beta^2 = c^2\left(\frac{\pi k}{2^n}\right)^2 = \left(\frac{\pi c\,k}{2^n}\right)^2 > 4\left(\frac{\pi|z|}{2^n}\right)^2 = 4|\alpha|^2 \quad \Longrightarrow \quad c^2\beta^2 - 4|\alpha|^2 > 0,$$

and combining this with the preceding line, we obtain

$$0 < c^2\beta^2 - 4|\alpha|^2 \leq |\sin^2 \alpha - \sin^2 \beta|.$$

Hence,

$$\frac{|\sin 2\alpha|}{|\sin^2 \alpha - \sin^2 \beta|} \leq \frac{3|\alpha|}{c^2\beta^2 - 4|\alpha|^2} = \frac{3\pi|z|/2^n}{c^2(\pi k/2^n)^2 - 4(\pi|z|/2^n)^2} = \frac{2^n}{\pi}\frac{3|z|}{c^2k^2 - 4|z|^2}.$$

Thus, for $n > N$ and $k \geq m$, in view of (7.18) and the definition of $a_k(n)$, we have

$$|a_k(n)| \leq M_k\ , \quad \text{where} \quad M_k = \frac{3|z|^2}{c^2k^2 - 4|z|^2}.$$

Since

$$M_k = \frac{3|z|^2}{c^2 - \frac{4|z|^2}{k^2}} \cdot \frac{1}{k^2} \leq \frac{3|z|^2}{c^2 - \frac{4|z|^2}{m^2}} \cdot \frac{1}{k^2} = \big(\text{constant}\big) \cdot \frac{1}{k^2},$$

by the comparison test, the sum $\sum_{k=m}^{\infty} M_k$ converges. This completes the proof of Euler's cotangent expansion.

7.4.2. Partial fraction expansions of the other trig functions.

We shall leave most of the details to the exercises. Using the formula (see (7.15))

$$\pi \tan \frac{\pi z}{2} = \pi \cot \frac{\pi z}{2} - 2\pi \cot \pi z,$$

and substituting in the partial fraction expansion of the cotangent, gives, as the diligent reader will do in Problem 1, for $z \in \mathbb{C}$ not an odd integer,

$$\boxed{\pi \tan \frac{\pi z}{2} = \sum_{n=0}^{\infty} \frac{4z}{(2n+1)^2 - z^2}.} \tag{7.19}$$

To derive a partial fraction expansion for $\frac{\pi}{\sin \pi z}$, we first derive the identity

$$\frac{1}{\sin z} = \cot z + \tan \frac{z}{2}.$$

To see this, observe that

$$\begin{aligned} \cot z + \tan \frac{z}{2} = \frac{\cos z}{\sin z} + \frac{\sin(z/2)}{\cos(z/2)} &= \frac{\cos z \, \cos(z/2) + \sin z \, \sin(z/2)}{\sin z \, \cos(z/2)} \\ &= \frac{\cos(z - (z/2))}{\sin z \, \cos(z/2)} = \frac{\cos(z/2)}{\sin z \, \cos(z/2)} = \frac{1}{\sin z}. \end{aligned}$$

This identity, together with the partial fraction expansions of the tangent and cotangent and a little algebra, which the extremely diligent reader will supply in Problem 1, imply that for noninteger $z \in \mathbb{C}$,

$$\boxed{\frac{\pi}{\sin \pi z} = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{n^2 - z^2}.} \tag{7.20}$$

Finally, the incredibly awesome diligent reader ☺ will supply the details for the following cosine expansion: For $z \in \mathbb{C}$ not an odd integer,

$$\boxed{\frac{\pi}{4 \cos \frac{\pi z}{2}} = \sum_{n=0}^{\infty} (-1)^n \frac{(2n+1)}{(2n+1)^2 - z^2}.} \tag{7.21}$$

Exercises 7.4.

1. Fill in the details for the proofs of (7.19) and (7.20). For (7.21), first show that
$$\frac{\pi}{\sin \pi z} = \frac{1}{z} + \Big(\frac{1}{1-z} - \frac{1}{1+z}\Big) - \Big(\frac{1}{2-z} - \frac{1}{2+z}\Big) + \cdots .$$
Replacing z with $\frac{1-z}{2}$ and doing some algebra, derive the expansion (7.21).
2. Derive Gregory-Leibniz-Madhava's series $\frac{\pi}{4} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} = 1 - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \cdots$ by replacing $z = 1/4$ in the partial fraction expansions of $\pi z \cot \pi z$ and $\pi / \sin \pi z$. How can you derive Gregory-Leibniz-Madhava's series from the expansion of $\frac{\pi}{4 \cos \frac{\pi z}{2}}$?
3. Derive the following formulas for π:
$$\pi = z \tan\Big(\frac{\pi}{z}\Big) \cdot \Big[1 - \frac{1}{z-1} + \frac{1}{z+1} - \frac{1}{2z-1} + \frac{1}{2z+1} - + \cdots\Big]$$
and
$$\pi = z \sin\Big(\frac{\pi}{z}\Big) \cdot \Big[1 + \frac{1}{z-1} - \frac{1}{z+1} - \frac{1}{2z-1} + \frac{1}{2z+1} + - - + + \cdots\Big].$$
In particular, plug in $z = 3, 4, 6$ to derive some pretty formulas.

7.5. ★ More proofs that $\pi^2/6 = \sum_{n=1}^\infty 1/n^2$

In this section, we continue our discussion from Sections 5.2 and 6.11, concerning the Basel problem of determining the sum of the reciprocals of the squares. A good reference for this material is [**109**] and for more on Euler, see [**11**].

7.5.1. Proof VIII of Euler's formula for $\pi^2/6$. (Cf. [**47**, p. 74].) One can consider this proof as a "logarithmic" version of Euler's original (third) proof of the formula for $\pi^2/6$, which we explained in the introduction to Chapter 5. As with Euler, we begin with Euler's sine expansion restricted to $0 \le x < 1$:

$$\frac{\sin \pi x}{\pi x} = \prod_{n=1}^\infty \Big(1 - \frac{x^2}{n^2}\Big).$$

However, in contrast to Euler, we take logarithms of both sides:

$$\begin{aligned}\log\left(\frac{\sin \pi x}{\pi x}\right) = \log\left(\lim_{m\to\infty} \prod_{n=1}^m \Big(1 - \frac{x^2}{n^2}\Big)\right) &= \lim_{m\to\infty} \log\left(\prod_{n=1}^m \Big(1 - \frac{x^2}{n^2}\Big)\right)\\ &= \lim_{m\to\infty} \sum_{n=1}^m \log\Big(1 - \frac{x^2}{n^2}\Big),\end{aligned}$$

where in the second equality we can pull out the limit because log is continuous, and at the last step we used that logarithms take products to sums. Thus, we have shown that

$$\log\left(\frac{\sin \pi x}{\pi x}\right) = \sum_{n=1}^\infty \log\Big(1 - \frac{x^2}{n^2}\Big), \quad 0 \le x < 1.$$

Recalling that $\log(1+t) = \sum_{m=1}^\infty \frac{(-1)^{m-1}}{m} t^m$, we see that

$$\log(1-t) = -\sum_{m=1}^\infty \frac{1}{m} t^m,$$

so replacing t by x^2/n^2 we obtain

$$\log\left(\frac{\sin \pi x}{\pi x}\right) = -\sum_{n=1}^\infty \sum_{m=1}^\infty \frac{1}{m} \frac{x^{2m}}{n^{2m}}, \quad 0 \le x < 1.$$

Since

$$\sum_{n=1}^\infty \sum_{m=1}^\infty \left|\frac{1}{m} \frac{x^{2m}}{n^{2m}}\right| = \sum_{n=1}^\infty \sum_{m=1}^\infty \frac{1}{m} \frac{|x|^{2m}}{n^{2m}} = \sum_{n=1}^\infty \log\Big(1 - \frac{|x|^2}{n^2}\Big) = \log\left(\frac{\sin \pi |x|}{\pi |x|}\right) < \infty,$$

by Cauchy's double series theorem, we can iterate sums:

$$\begin{aligned}(7.22) \qquad -\log\left(\frac{\sin \pi x}{\pi x}\right) &= \sum_{m=1}^\infty \left(\sum_{n=1}^\infty \frac{1}{n^{2m}}\right) \frac{x^{2m}}{m}\\ &= x^2 \sum_{n=1}^\infty \frac{1}{n^2} + \frac{x^4}{2} \sum_{n=1}^\infty \frac{1}{n^4} + \frac{x^6}{3} \sum_{n=1}^\infty \frac{1}{n^6} + \cdots .\end{aligned}$$

On the other hand, by our power series composition theorem, we have (after some simplification)

$$-\log\left(\frac{\sin\pi x}{\pi x}\right) = -\log\left(1 - \left(\frac{\pi^2x^2}{3!} - \frac{\pi^4x^4}{5!} + - \cdots\right)\right)$$
$$= \left(\frac{\pi^2x^2}{3!} - \frac{\pi^4x^4}{5!} + - \cdots\right) + \frac{1}{2}\left(\frac{\pi^2x^2}{3!} - \frac{\pi^4x^4}{5!} + - \cdots\right)^2 + \cdots$$
$$(7.23)\qquad = \frac{\pi^2}{3!}x^2 + \left(-\frac{\pi^4}{5!} + \frac{\pi^4}{2\cdot(3!)^2}\right)x^4 + \left(\frac{\pi^6}{7!} - \frac{\pi^6}{3!\cdot 5!} + \frac{\pi^6}{3\cdot(3!)^3}\right)x^6 + \cdots .$$

Equating this with (7.22), we obtain

$$\frac{\pi^2}{3!}x^2 + \left(-\frac{\pi^4}{5!} + \frac{\pi^4}{2\cdot(3!)^2}\right)x^4 + \left(\frac{\pi^6}{7!} - \frac{\pi^6}{3!\cdot 5!} + \frac{\pi^6}{3\cdot(3!)^3}\right)x^6 + \cdots$$
$$= x^2\sum_{n=1}^\infty \frac{1}{n^2} + \frac{x^4}{2}\sum_{n=1}^\infty \frac{1}{n^4} + \frac{x^6}{3}\sum_{n=1}^\infty \frac{1}{n^6} + \cdots ,$$

or after simplification,

$$(7.24)\quad \frac{\pi^2}{6}x^2 + \frac{\pi^4}{180}x^4 + \frac{\pi^6}{2835}x^6 + \cdots = x^2\sum_{n=1}^\infty \frac{1}{n^2} + \frac{x^4}{2}\sum_{n=1}^\infty \frac{1}{n^4} + \frac{x^6}{3}\sum_{n=1}^\infty \frac{1}{n^6} + \cdots .$$

By the identity theorem, the coefficients of x^k must be identical. Thus, comparing the x^2 terms, we get Euler's formula:

$$\frac{\pi^2}{6} = \sum_{n=1}^\infty \frac{1}{n^2},$$

comparing the x^4 terms, we get

$$(7.25)\qquad \boxed{\frac{\pi^4}{90} = \sum_{n=1}^\infty \frac{1}{n^4},}$$

and finally, comparing the x^6 terms, we get

$$(7.26)\qquad \boxed{\frac{\pi^6}{945} = \sum_{n=1}^\infty \frac{1}{n^6}.}$$

Now what if we took more terms in (7.22) and (7.23), say to x^{2k}, can we then find a formula for $\sum 1/n^{2k}$? The answer is certainly true but the work required to get a formula is rather intimidating; see Problem 1 for a formula when $k = 4$. Of course, in Section 5.2 we found formulas for $\zeta(2k)$ for *all* k.

7.5.2. Proof IX. (Cf. [**123**], [**49**].) For this proof, we start with Lemma 7.8, which states that if $n = 2m + 1$ with $m \in \mathbb{N}$, then

$$(7.27)\qquad \sin nz = n\sin z\prod_{k=1}^m\left(1 - \frac{\sin^2 z}{\sin^2(k\pi/n)}\right).$$

We fix an m; later we shall take $m \to \infty$. We now substitute the expansion

$$\sin nz = nz - \frac{n^3z^3}{3!} + \frac{n^5z^5}{5!} - + \cdots$$

into the left-hand side of (7.27), and the expansions

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - + \cdots ,$$

and

$$\sin^2 z = \frac{1}{2}(1 - \cos 2z) = z^2 - \frac{2}{3}z^4 + - \cdots ,$$

into the right-hand side of (7.27). Then multiplying everything out and simplifying, we obtain (after a lot of algebra)

$$nz - \frac{n^3\, z^3}{3!} + - \cdots = nz + \left(-\frac{n}{6} - n \sum_{k=1}^{m} \frac{1}{\sin^2(k\pi/n)} \right) z^3 + \cdots .$$

Comparing the z^3 terms, by the identity theorem we conclude that

$$-\frac{n^3}{6} = -\frac{n}{6} - n \sum_{k=1}^{m} \frac{1}{\sin^2(k\pi/n)},$$

which can be written in the form

$$\frac{1}{6} - \sum_{k=1}^{m} \frac{1}{n^2 \sin^2(k\pi/n)} = \frac{1}{6n^2}. \tag{7.28}$$

To establish Euler's formula, we apply Tannery's theorem to this sum. According to Lemma 5.6, for some positive constant c,

$$c\, x \le \sin x \qquad \text{for } 0 \le x \le \pi/2. \tag{7.29}$$

Now for $0 \le k \le m = (n-1)/2$, we have $k\pi/n < \pi/2$, so for such k,

$$c \cdot \frac{k\pi}{n} \le \sin \frac{k\pi}{n},$$

which gives

$$\frac{1}{n^2} \cdot \frac{1}{\sin^2(k\pi/n)} \le \frac{1}{n^2} \cdot \frac{n^2}{(c\pi)^2 k^2} = \frac{1}{c^2\pi^2} \cdot \frac{1}{k^2}.$$

By the p-test, we know that the sum

$$\sum_{k=1}^{\infty} \frac{1}{c^2\pi^2} \cdot \frac{1}{k^2}$$

converges. Also, since $n \sin(x/n) \to x$ as $n \to \infty$, which implies that

$$\lim_{n\to\infty} \frac{1}{n^2 \sin^2(k\pi/n)} = \frac{1}{k^2\pi^2},$$

taking $m \to \infty$ in (7.28), Tannery's theorem gives

$$\frac{1}{6} - \sum_{k=1}^{\infty} \frac{1}{k^2\pi^2} = 0,$$

which is equivalent to Euler's formula. See Problem 2 for a proof that uses (7.28) but doesn't use Tannery's theorem.

Exercises 7.5.

1. Determine the sum $\sum_{n=1}^{\infty} \frac{1}{n^8}$ using Euler's method; that is, in the same manner as we derived (7.25) and (7.26).
2. (Cf. [**49**]) (**Euler's sum, Proof X**) Instead of using Tannery's theorem to derive Euler's formula from (7.28), we can follow Kortram [**123**] as follows.

(i) Fix any $M \in \mathbb{N}$ and let $m > M$. Using (7.28), prove that for $n = 2m+1$,

$$\frac{1}{6} - \sum_{k=1}^{M} \frac{1}{n^2 \sin^2(k\pi/n)} = \frac{1}{n^2} + \sum_{k=M+1}^{m} \frac{1}{n^2 \sin^2(k\pi/n)}.$$

(ii) Using that $cx \le \sin x$ for $0 \le x \le \pi/2$ with $c > 0$, prove that

$$0 \le \frac{1}{6} - \sum_{k=1}^{M} \frac{1}{n^2 \sin^2(k\pi/n)} \le \frac{1}{n^2} + \frac{1}{c^2\pi^2} \sum_{k=M+1}^{\infty} \frac{1}{k^2}.$$

(iii) Finally, letting $m \to \infty$ (so that $n = 2m+1 \to \infty$ as well) and then letting $M \to \infty$, establish Euler's formula.

3. (Cf. [**56**]) Let $S \subseteq \mathbb{N}$ denote the set of square-free natural numbers; see Subsection 6.7.2 for a review of square-free numbers.

(i) Let $N \in \mathbb{N}$ and prove that

$$\sum_{n<N} \frac{1}{n^2} \le \Big(\sum_{k<N} \frac{1}{k^4} \Big) \Big(\sum_{n \in S\,,\, n<N} \frac{1}{n^2} \Big) \le \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

(ii) If $\sum_{n\in S} \frac{1}{n^2} := \lim_{N\to\infty} \sum_{n\in S\,,\,n<N} \frac{1}{n^2}$, using (i), prove that

$$\sum_{n\in S} \frac{1}{n^2} = \frac{15}{\pi^2}.$$

4. (Cf. [**56**]) Let $A \subseteq \mathbb{N}$ denote the set of natural numbers that are not perfect squares. With $\sum_{n\in A} \frac{1}{n^2} := \lim_{N\to\infty} \sum_{n\in A\,,\,n<N} \frac{1}{n^2}$, prove that

$$\sum_{n\in A} \frac{1}{n^2} = \frac{\pi^2}{90}(15 - \pi^2).$$

7.6. ★ Riemann's remarkable ζ-function, probability, and $\pi^2/6$

We have already seen the Riemann zeta function at work in many examples. In this section we're going to look at some of its relations with number theory; this will give just a hint as to the great importance of the zeta function in mathematics. As a consolation prize to our discussion on Riemann's ζ-function we'll find an incredible connection between probability theory and $\pi^2/6$.

7.6.1. The Riemann-zeta function and number theory. We begin with the following theorem proved by Euler which connects $\zeta(z)$ to prime numbers. See Problem 1 for a proof of this theorem using the good ole Tannery's theorem!

Theorem 7.12 (**Euler and Riemann**). *For all $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$, we have*

$$\boxed{\zeta(z) = \prod \Big(1 - \frac{1}{p^z}\Big)^{-1} = \prod \frac{p^z}{p^z - 1},}$$

where the infinite product is over all prime numbers $p \in \mathbb{N}$.

Proof. We give two proofs, one using Cauchy's theorem on the multiplication of series and the other is Euler's classic.

Proof I: Let $r > 1$ be arbitrary and let $\operatorname{Re} z \ge r$. Let $2 < N \in \mathbb{N}$ and let $2 < 3 < \cdots < m < N$ be all the primes less than N. Then for every natural $n < N$, by unique factorization,

$$n^z = \big(2^i\, 3^j\, \cdots\, m^k\big)^z = 2^{iz}\, 3^{jz}\, \cdots\, m^{kz}$$

for some nonnegative integers $i, j, \dots, k$. Using this fact, it follows that the product

$$\begin{aligned}\prod_{p<N}\Big(1-\frac{1}{p^z}\Big)^{-1} &= \Big(1-\frac{1}{2^z}\Big)^{-1}\Big(1-\frac{1}{3^z}\Big)^{-1}\cdots\Big(1-\frac{1}{m^z}\Big)^{-1}\\ &= \Big(1+\frac{1}{2^z}+\frac{1}{2^{2z}}+\frac{1}{2^{3z}}\cdots\Big)\Big(1+\frac{1}{3^z}+\frac{1}{3^{2z}}+\frac{1}{3^{3z}}+\cdots\Big)\cdots\\ &\qquad\cdots\Big(1+\frac{1}{m^z}+\frac{1}{m^{2z}}+\frac{1}{m^{3z}}+\cdots\Big),\end{aligned}$$

after multiplying out and using Cauchy's multiplication theorem (or rather its generalization to a product of more than two absolutely convergent series), contains the numbers $1, \frac{1}{2^z}, \frac{1}{3^z}, \frac{1}{4^z}, \frac{1}{5^z} \dots, \frac{1}{(N-1)^z}$ (along with all other numbers $\frac{1}{n^z}$ with $n \geq N$ having prime factors $2, 3, \dots, m$). In particular,

$$\Big|\sum_{n=1}^{\infty}\frac{1}{n^z}-\prod_{p<N}\Big(1-\frac{1}{p^z}\Big)^{-1}\Big| \leq \sum_{n=N}^{\infty}\Big|\frac{1}{n^z}\Big| \leq \sum_{n=N}^{\infty}\frac{1}{n^r},$$

since $\operatorname{Re} z \geq r$. By the p-test (with $p = r > 1$), $\sum \frac{1}{n^r}$ converges so the right-hand side tends to zero as $N \to \infty$. This completes Proof I.

Proof II: Here's Euler's beautiful proof using a "sieving method" made famous by Eratosthenes of Cyrene (276 B.C.–194 B.C.). First we get rid of all the numbers in $\zeta(z)$ that have factors of 2: Observe that

$$\frac{1}{2^z}\zeta(z) = \frac{1}{2^z}\sum_{n=1}^{\infty}\frac{1}{n^z} = \sum_{n=1}^{\infty}\frac{1}{(2n)^z},$$

therefore,

$$\Big(1-\frac{1}{2^z}\Big)\zeta(z) = \sum_{n=1}^{\infty}\frac{1}{n^z} - \sum_{n=1}^{\infty}\frac{1}{(2n)^z} = \sum_{n\,;\,2\not| n}\frac{1}{n^z}.$$

Next, we get rid of all the numbers in $\big(1-\frac{1}{2^z}\big)\zeta(z)$ that have factors of 3: Observe that

$$\frac{1}{3^z}\Big(1-\frac{1}{2^z}\Big)\zeta(z) = \frac{1}{3^z}\sum_{2\not| n}\frac{1}{n^z} = \sum_{n\,;\,2\not| n}\frac{1}{(3n)^z},$$

therefore,

$$\begin{aligned}\Big(1-\frac{1}{3^z}\Big)\Big(1-\frac{1}{2^z}\Big)\zeta(z) &= \Big(1-\frac{1}{2^z}\Big)\zeta(z) - \frac{1}{3^z}\Big(1-\frac{1}{2^z}\Big)\zeta(z)\\ &= \sum_{n\,;\,2\not| n}\frac{1}{n^z} - \sum_{n\,;\,2\not| n}\frac{1}{(3n)^z}\\ &= \sum_{n\,;\,2,3\not| n}\frac{1}{n^z}.\end{aligned}$$

Repeating this argument, we get, for any prime q:

$$\Big\{\prod_{p\ \text{prime}\leq q}\Big(1-\frac{1}{p^z}\Big)\Big\}\zeta(z) = \sum_{n\,;\,2,3,\dots,q\not| n}\frac{1}{n^z},$$

where the sum is over all $n \in \mathbb{N}$ that are not divisible by the primes from 2 to q. Therefore, choosing $r > 1$ such that $|z| > r$, we have

$$\left|\left\{\prod_{p \text{ prime}\leq q}\left(1 - \frac{1}{p^z}\right)\right\}\zeta(z) - 1\right| = \left|\sum_{n\,;\,n\neq 1\ \&\ 2,3,\ldots,q \not{|} n} \frac{1}{n^z}\right|$$
$$\leq \sum_{n\,;\,n\neq 1\ \&\ 2,3,\ldots,q \not{|} n} \frac{1}{n^r} \leq \sum_{n=q}^{\infty} \frac{1}{n^r}.$$

By Cauchy's criterion for series, $\lim_{q\to\infty} \sum_{n=q}^{\infty} \frac{1}{n^r} = 0$, so we conclude that

$$\left\{\prod_{p \text{ prime}}\left(1 - \frac{1}{p^z}\right)\right\}\zeta(z) = 1,$$

which is equivalent to Euler's product formula. □

In particular, since we know that $\zeta(2) = \pi^2/6$, we have

$$\boxed{\frac{\pi^2}{6} = \prod \frac{p^2}{p^2 - 1} = \frac{2^2}{2^2 - 1} \cdot \frac{3^2}{3^2 - 1} \cdot \frac{5^2}{5^2 - 1} \cdots}.$$

Our next connection is with the following strange (but interesting) function:

$$\mu(n) := \begin{cases} 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n = p_1\, p_2\, \cdots p_k \text{ is a product } k \text{ distinct prime numbers} \\ 0 & \text{else.} \end{cases}$$

This function is called the **Möbius function** after August Ferdinand Möbius (1790–1868) who introduced the function in 1831. Some of its values are

$$\mu(1) = 1\ ,\ \mu(2) = -1\ ,\ \mu(3) = -1\ ,\ \mu(4) = 0\ ,\ \mu(5) = -1\ ,\ \mu(6) = 1\ ,\ \ldots.$$

THEOREM 7.13. *For all $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$, we have*

$$\boxed{\frac{1}{\zeta(z)} = \prod\left(1 - \frac{1}{p^z}\right) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^z}.}$$

PROOF. Let $r > 1$ be arbitrary and let $\operatorname{Re} z \geq r$. Let $2 < N \in \mathbb{N}$ and let $2 < 3 < \cdots < m < N$ be all the primes less than N. Then observe that the product

$$\prod_{n<N}\left(1 - \frac{1}{p^z}\right) = \left(1 + \frac{-1}{2^z}\right)\left(1 + \frac{-1}{3^z}\right)\left(1 + \frac{-1}{5^z}\right)\cdots\left(1 + \frac{-1}{m^z}\right),$$

when multiplied out contains 1 and all numbers of the form

$$\left(\frac{-1}{p_1^z}\right)\cdot\left(\frac{-1}{p_2^z}\right)\cdot\left(\frac{-1}{p_3^z}\right)\cdots\left(\frac{-1}{p_k^z}\right) = \frac{(-1)^k}{p_1^z p_2^z \cdots p_k^z} = \frac{(-1)^k}{n^z}, \qquad n = p_1\, p_2\, \ldots\, p_k,$$

where $p_1 < p_2 < \cdots < p_k < N$ are distinct primes. In particular, $\prod_{n<N}\left(1 - \frac{1}{p^z}\right)$ contains the numbers $\frac{\mu(n)}{n^z}$ for $n = 1, 2, \ldots, N-1$ (along with all other numbers $\frac{\mu(n)}{n^z}$ with $n \geq N$ having prime factors $2, 3, \ldots, m$), so

$$\left|\sum_{n=1}^{\infty} \frac{\mu(n)}{n^z} - \prod_{p<N}\left(1 - \frac{1}{p^z}\right)\right| \leq \sum_{n=N}^{\infty}\left|\frac{\mu(n)}{n^z}\right| \leq \sum_{n=N}^{\infty} \frac{1}{n^r},$$

since $\operatorname{Re} z \geq r$. By the p-test (with $p = r > 1$), $\sum \frac{1}{n^r}$ converges so the right-hand side tends to zero as $N \to \infty$. This completes our proof. □

See the exercises for other neat connections of $\zeta(z)$ with number theory.

7.6.2. The eta function. A function related to the zeta function is the "alternating zeta function" or **Dirichlet eta-function**:

$$\eta(z) := \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^z}.$$

We can write the eta function in terms of the zeta function as follows.

Theorem 7.14. *We have*

$$\eta(z) = (1 - 2^{1-z})\zeta(z)\,, \quad z > 1.$$

Proof. Splitting into sums of even and odd numbers, we get

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^z} &= -\sum_{n=1}^{\infty} \frac{1}{(2n)^z} + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^z} \\ &= -\sum_{n=1}^{\infty} \frac{1}{2^z}\frac{1}{n^z} + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^z} \\ &= -2^{-z}\zeta(z) + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^z}.\end{aligned}$$

On the other hand, breaking the zeta function into sums of even and odd numbers, we get

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z} = \sum_{n=1}^{\infty} \frac{1}{(2n)^z} + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^z} = 2^{-z}\zeta(z) + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^z}.$$

Substituting this expression into the previous one, we see that

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^z} = -2^{-z}\zeta(z) + \zeta(z) - 2^{-z}\zeta(z),$$

which is equivalent to the expression that we desired to prove. □

We now consider a shocking connection between probability theory, prime numbers, divisibility, and $\pi^2/6$ (cf. [**2**], [**107**]).[1] **Question:** What is the probability that a natural number, chosen at random, is square free? Answer (drum role please): $6/\pi^2$, a result which follows from work of Dirichlet in 1849 [**122**, p. 324], [**95**, p. 272]. Here's another **Question:** What is the probability that two given numbers, chosen at random, are relatively prime? Answer (drum role please): $6/\pi^2$, first proved by Leopold Bernhard Gegenbauer (Feb 1849–1903) [**95**, p. 272] who proved it in 1885.

[1] *Such shocking connections in science perhaps made Albert Einstein (1879–1955) state that "the scientist's religious feeling takes the form of a rapturous amazement at the harmony of natural law, which reveals an intelligence of such superiority that, compared with it, all the systematic thinking and acting of human beings is an utterly insignificant reflection".* [**103**]

7.6.3. Elementary probability theory. You will prove these results with complete rigor in Problems 11 and 10. However, we are going to derive them intuitively — *not* rigorously (!) — based on some basic probability ideas that should be "obvious" (or at least believable) to you; see [**229, 70, 71**] for standard books on probability in case you want the hardcore theory. We only need the basics. We denote the probability, or chance, that an event A happens by $P(A)$. The classic definition is

$$P(A) = \frac{\text{number of occurrences of } A}{\text{total number of possibilities}}. \tag{7.30}$$

For example, consider a classroom with 10 people, m men and w women (so that $m + w = 10$). The probability of randomly "choosing a man" ($= M$) is

$$P(M) = \frac{\text{number of men}}{\text{total number of possibilities}} = \frac{m}{10}.$$

Similarly, the probability of randomly choosing a woman is $w/10$. We next need to discuss complementary events. If A^c is the event that A does not happen, then

$$P(A^c) = 1 - P(A). \tag{7.31}$$

For instance, according to (7.31) the probability of "*not* choosing a man", M^c, should be $P(M^c) = 1 - P(M) = 1 - m/10$. But this is certainly true because "*not* choosing a man" is the same as "choosing a woman" W, so recalling that $m + w = 10$, we have

$$P(M^c) = P(W) = \frac{w}{10} = \frac{10 - m}{10} = 1 - \frac{m}{10}.$$

Finally, we need to discuss independence. Whenever an event A is *unrelated* to an event B (such events are called **independent**), we have the fundamental relation:

$$P(A \text{ and } B) = P(A) \cdot P(B).$$

For example, let's say that we have two classrooms of 10 students each, the first one with m_1 men and w_1 women, and the second one with m_2 men and w_2 women. Let us randomly choose a pair of students, one from the first classroom and the other from the second. What is the probability of "choosing a man from the first classroom" $= A$ *and* "choosing a woman from the second classroom" $= B$? Certainly A and B don't depend on each other, so by our formula above we should have

$$P(A \text{ and } B) = P(A) \cdot P(B) = \frac{m_1}{10} \cdot \frac{w_2}{10} = \frac{m_1 w_2}{100}.$$

To see that this is indeed true, note that the number of ways to pair a man in classroom 1 with a woman in classroom 2 is $m_1 \cdot w_2$ and the total number of possible pairs of people is $10^2 = 100$. Thus,

$$P(A \text{ and } B) = \frac{\text{number of men-women pairs}}{\text{total number of possible pairs of people}} = \frac{m_1 \cdot m_2}{100},$$

in agreement with our previous calculation. We remark that for any number of events $A_1, A_2, \ldots$, which are unrelated to each other, we have the generalized result:

$$P(A_1 \text{ and } A_2 \text{ and } \cdots) = P(A_1) \cdot P(A_2) \cdots . \tag{7.32}$$

7.6.4. Probability and $\pi^2/6$. To begin discussing our two incredible and shocking problems, we first look at the following question: Given a natural number k, what is the probability, or chance, that a randomly chosen natural number is divisible by k? Since the definition (7.30) involves finite quantities, we can't use this definition as it stands. We can instead use the following modified version:

$$P(A) = \lim_{n\to\infty} \frac{\text{number of occurrences of } A \text{ amongst } n \text{ possibilities}}{n}. \tag{7.33}$$

Using this formula, in Problem 8, you should be able to prove that the probability a randomly chosen natural number is divisible by k is $1/k$. However, instead of using (7.33), we shall employ the following heuristic trick (which works to give the correct answer). Choose an "extremely large" natural number N, and consider the very large sample of numbers

$$1, 2, 3, 4, 5, 6, \ldots, Nk.$$

There are exactly N numbers in this list that are divisible by k, namely the N numbers $k, 2k, 3k, \ldots, Nk$, and no others, and there are a total of Nk numbers in this list. Thus, the probability that a natural number n, randomly chosen amongst the large sample, is divisible by k is exactly the probability that n is one of the N numbers $k, 2k, 3k, \ldots, Nk$, so

$$P(k|n) = \frac{\text{number of occurrences of divisibility}}{\text{total number of possibilities listed}} = \frac{N}{Nk} = \frac{1}{k}. \tag{7.34}$$

For instance, the probability that a randomly chosen natural number is divisible by 1 is 1, which makes sense. The probability that a randomly chosen natural number is divisible by 2 is $1/2$; in other words, the probability that a randomly chosen natural number is even is $1/2$, which also makes sense.

We are now ready to solve our two problems. **Question:** What is the probability that a natural number, chosen at random, is square free? Let $n \in \mathbb{N}$ be randomly chosen. Then n is square free just means that $p^2 \not| n$ for all primes p. Thus,

$$P(n \text{ is square free}) = P((2^2 \not| n) \text{ and } (3^2 \not| n) \text{ and } (5^2 \not| n) \text{ and } (7^2 \not| n) \text{ and } \cdots).$$

Since n was randomly chosen, the events $2^2 \not| n$, $3^2 \not| n$, $5^2 \not| n$, etc. are unrelated, so by (7.32),

$$P(n \text{ is square free}) = P(2^2 \not| n) \cdot P(3^2 \not| n) \cdot P(5^2 \not| n) \cdot P(7^2 \not| n) \cdots$$

To see what the right-hand side is, we use (7.31) and (7.34) to write

$$P(p^2 \not| n) = 1 - P(p^2 | n) = 1 - \frac{1}{p^2}.$$

Thus,

$$P(n \text{ is square free}) = \prod_{p \text{ prime}} P(p^2 \not| n) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2},$$

and our first question is answered!

Question: What is the probability that two given numbers, chosen at random, are relatively, or co, prime? Let $m, n \in \mathbb{N}$ be randomly chosen. Then m and n are

relatively prime, or **coprime**, just means that m and n have no common factors (except 1), which means[2] that $p \nmid$ both m, n for all prime numbers p. Thus,

$$\begin{aligned}&P(m, n \text{ are relatively prime})\\ &\qquad = P((2 \nmid \text{both } m, n) \text{ and } (3 \nmid \text{both } m, n) \text{ and } (5 \nmid \text{both } m, n) \text{ and } \cdots).\end{aligned}$$

Since m and n were randomly chosen, that $p \nmid$ both m, n is unrelated to $q \nmid$ both m, n, so by (7.32),

$$P(m, n \text{ are relatively prime}) = \prod_{p \text{ prime}} P(p \nmid \text{both } m, n).$$

To see what the right-hand side is, we use (7.31), (7.32), and (7.34) to write

$$\begin{aligned}P(p \nmid \text{both } m, n) &= 1 - P(p| \text{both } m, n) = 1 - P(p|m \text{ and } p|n)\\ &= 1 - P(p|m) \cdot P(p|n) = 1 - \frac{1}{p} \cdot \frac{1}{p} = 1 - \frac{1}{p^2}.\end{aligned}$$

Thus,

$$P(m, n \text{ are relatively prime}) = \prod_{p \text{ prime}} P(p \nmid \text{both } m, n) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^2}\right) = \frac{6}{\pi^2},$$

and our second question is answered!

Exercises 7.6.

1. ($\zeta(z)$ **product formula, Proof III**) We prove Theorem 7.12 using the good ole Tannery's theorem for products.
   (i) Let $r > 1$ be arbitrary and let $\operatorname{Re} z \geq r$. Prove that
   $$\left| \prod_{p<N} \frac{p^z - (1/p^z)^N}{p^z - 1} - \sum_{n=1}^{\infty} \frac{1}{n^z} \right| \leq \sum_{n=N}^{\infty} \frac{1}{n^r}.$$
   Suggestion: $\frac{p^z-(1/p^z)^N}{p^z-1} = \frac{1-(1/p^z)^{N+1}}{1-1/p^z} = 1 + 1/p^z + 1/p^{2z} + \cdots + 1/p^{Nz}$.
   (ii) Write $\frac{p^z-(1/p^z)^N}{p^z-1} = 1 + \frac{1-(1/p^z)^N}{p^z-1}$. Show that
   $$\left| \frac{1-(1/p^z)^N}{p^z - 1} \right| \leq \frac{2}{p^r - 1} \leq \frac{4}{p^r}$$
   and $\sum 4/p^r$ converges. Now prove Theorem 7.12 using Tannery's theorem for products.
2. Prove that for $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$,
   $$\boxed{\frac{\zeta(z)}{\zeta(2z)} = \sum_{n=1}^{\infty} \frac{|\mu(n)|}{n^z}.}$$
   Suggestion: Show that $\frac{\zeta(z)}{\zeta(2z)} = \prod \left(1 + \frac{1}{p^z}\right)$ and copy Proof I of Theorem 7.13.
3. (**Möbius inversion formula**) In this problem we prove Möbius inversion formula.
   (i) Given $n \in \mathbb{N}$ with $n > 1$, let $p_1, \ldots, p_k$ be the distinct prime factors of n. For $1 \leq i \leq k$, let
   $$A_i = \{m \in \mathbb{N}\,;\, m = \text{a product of exactly } i \text{ distinct prime factors of } n\}.$$

[2] Explicitly, "$p \nmid$ both m, n" is the negation of "$p|m$ and $p|n$"; that is, "$p \nmid m$ or $p \nmid n$".

Show that

$$\sum_{d|n} \mu(d) = 1 + \sum_{i=1}^{k} \sum_{m \in A_i} \mu(m),$$

where $\sum_{d|n} \mu(d)$ means to sum over all $d \in \mathbb{N}$ such that $d|n$. Next, show that

$$\sum_{m \in A_i} \mu(m) = (-1)^i \binom{k}{i}.$$

(ii) For any $n \in \mathbb{N}$, prove that

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

(iii) Let $f : (0, \infty) \to \mathbb{R}$ such that $f(x) = 0$ for $x < 1$, and define

$$g(x) = \sum_{n=1}^{\infty} f\left(\frac{x}{n}\right).$$

Note that $g(x) = 0$ for $x < 1$ and this infinite series is really only a finite sum since $f(x) = 0$ for $x < 1$; specifically, choosing any $N \in \mathbb{N}$ with $N \geq \lfloor x \rfloor$ (the greatest integer $\leq x$), we have $g(x) = \sum_{n=1}^{N} f(x/n)$. Prove that

$$f(x) = \sum_{n=1}^{\infty} \mu(n)\, g\left(\frac{x}{n}\right) \qquad (\textbf{Möbius inversion formula});$$

As before, this sum is really only finite. Suggestion: If you've not gotten anywhere after some time, let $S = \{(k, n) \in \mathbb{N} \times \mathbb{N}\,;\, n|k\}$ and consider the sum

$$\sum_{(k,n) \in S} \mu(n)\, f\left(\frac{x}{k}\right).$$

Write this sum as $\sum_{k=1}^{\infty} \sum_{n\,;\,n|k} \mu(n)\, f(x/k)$, then as $\sum_{n=1}^{\infty} \sum_{k\,;\,n|k} \mu(n)\, f(x/k)$ and simplify each iterated sum.

4. (**Liouville's function**) Define

$$\lambda(n) := \begin{cases} 1 & \text{if } n = 1 \\ 1 & \text{if the number of prime factors of } n\text{, counted with repetitions, is even} \\ -1 & \text{if the number of prime factors of } n\text{, counted with repetitions, is odd.} \end{cases}$$

This function is called **Liouville's function** after Joseph Liouville (1809–1882). Prove that for $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$,

$$\boxed{\frac{\zeta(2z)}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^z}.}$$

Suggestion: Show that $\frac{\zeta(2z)}{\zeta(z)} = \prod \left(1 + \frac{1}{p^z}\right)^{-1}$ and copy Proof I of Theorem 7.12.

5. For $n \in \mathbb{N}$, let $\tau(n)$ denote the number of positive divisors of n (that is, the number of positive integers that divide n). Prove that for $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$,

$$\boxed{\zeta(z)^2 = \sum_{n=1}^{\infty} \frac{\tau(n)}{n^z}.}$$

Suggestion: By absolute convergence, we can write $\zeta(z)^2 = \sum_{m,n} 1/(m \cdot n)^z$ where this double series can be summed in any way we wish. Use Theorem 6.25 with the set S_k given by $S_k = T_1 \cup \cdots \cup T_k$ where $T_k = \{(m, n) \in \mathbb{N} \times \mathbb{N}\,;\, m \cdot n = k\}$.

6. Let $\zeta(z, a) := \sum_{n=0}^{\infty}(n+a)^{-z}$ for $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$ and $a > 0$ — this function is called the **Hurwitz zeta function** after Adolf Hurwitz (1859–1919). Prove that
$$\sum_{m=1}^{k} \zeta\Big(z, \frac{m}{k}\Big) = k^z \zeta(z).$$
7. In this problem, we find useful bounds and limits for $\zeta(x)$ with $x > 1$ real.
   (a) Prove that $1 - \frac{1}{2^x} < \eta(x) < 1$.
   (b) Prove that
   $$\frac{1-2^{-x}}{1-2^{1-x}} < \zeta(x) < \frac{1}{1-2^{1-x}}.$$
   (c) Prove the following limits: $\zeta(x) \to 1$ as $x \to \infty$, $\zeta(x) \to \infty$ as $x \to 1^+$, and $(x-1)\zeta(x) \to 1$ as $x \to 1^+$.
8. Using the definition (7.33), prove that given a natural number k, the probability that a randomly chosen natural number is divisible by k is $1/k$. Suggestion: Amongst the n natural numbers $1, 2, 3, \ldots, n$, show that $\lfloor n/k \rfloor$ many numbers are divisible by k. Now take $n \to \infty$ in $\lfloor n/k \rfloor / n$.
9. (cf. [**25, 107**]) Let $k \in \mathbb{N}$ with $k \geq 2$. We say that a natural number n is **k-th power free** if $p^k \nmid n$ for all primes p. What is the probability that a natural number, chosen at random, is k-th power free? What is the probability that k natural numbers, chosen at random, are relatively prime (have not common factors except 1)?
10. (**Square-free numbers**) Define $S : (0, \infty) \to \mathbb{R}$ by
$$S(x) := \#\{k \in \mathbb{N}\,;\, 1 \leq k \leq x \text{ and } k \text{ is square free}\};$$
note that $S(x) = 0$ for $x < 1$. We shall prove that
$$\lim_{n\to\infty} \frac{S(n)}{n} = \frac{6}{\pi^2}.$$
Do you see why this formula makes precise the statement "The probability that a randomly chosen natural number is square free equals $6/\pi^2$"?
   (i) For any real number $x > 0$ and $n \in \mathbb{N}$, define
   $$A(x, n) := \big\{k \in \mathbb{N}\,;\, 1 \leq k \leq x \text{ and } n^2 \text{ is the largest square that divides } k\big\}.$$
   Note that $A(x, n) = \varnothing$ for $n^2 > x$. Prove that $A(x, 1)$ consists of all square-free numbers $\leq x$, and also prove that
   $$\{k \in \mathbb{N}\,;\, 1 \leq k \leq x\} = \bigcup_{n=1}^{\infty} A(x, n).$$
   (ii) Show that there is a bijection between $A(x, n)$ and $A(x/n^2, 1)$.
   (iii) Show that for any $x > 0$, we have
   $$\lfloor x \rfloor = \sum_{n=1}^{\infty} S\Big(\frac{x}{n^2}\Big).$$
   Using the Möbius inversion formula from Problem 3, conclude that
   $$S(x) = \sum_{n=1}^{\infty} \mu(n) \left\lfloor \frac{x}{n^2} \right\rfloor.$$
   (iv) Finally, prove that $\lim_{x\to\infty} S(x)/x = 6/\pi^2$, which in particular proves our result.
11. (**Relatively prime numbers**; for different proofs, see [**122**, p. 337] and [**95**, p. 268]) Define $R : (0, \infty) \to \mathbb{R}$ by
$$R(x) := \#\{(k, \ell) \in \mathbb{N}\,;\, 1 \leq k, \ell \leq x \text{ and } k \text{ and } \ell \text{ are relatively prime}\};$$

note that $R(x) = 0$ for $x < 1$. We shall prove that

$$\lim_{n\to\infty} \frac{R(n)}{n^2} = \frac{6}{\pi^2},$$

Do you see why this formula makes precise the statement "The probability that two randomly chosen natural numbers are relatively prime equals $6/\pi^2$"?

(i) For any real number $x > 0$ and $n \in \mathbb{N}$, define

$$A(x,n) := \big\{(k,\ell) \in \mathbb{N} \times \mathbb{N}\,;\, 1 \le k, \ell \le x \text{ and } n \text{ is the largest divisor of both } k \text{ and } \ell\big\}.$$

Note that $A(x,n) = \varnothing$ for $n > x$. Prove that $A(x,1)$ consists of all pairs (k,ℓ) of relatively prime natural numbers that are $\le x$, and also prove that

$$\{(k,\ell) \in \mathbb{N} \times \mathbb{N}\,;\, 1 \le k, \ell \le x\} = \bigcup_{n=1}^{\infty} A(x,n).$$

(ii) Show that there is a bijection between $A(x,n)$ and $A(x/n,1)$.

(iii) Show that for any $x > 0$, we have

$$\lfloor x \rfloor^2 = \sum_{n=1}^{\infty} R\left(\frac{x}{n}\right).$$

Using the Möbius inversion formula from Problem 3, conclude that

$$R(x) = \sum_{n=1}^{\infty} \mu(n) \left\lfloor \frac{x}{n} \right\rfloor^2.$$

(iv) Finally, prove that $\lim_{x\to\infty} R(x)/x^2 = 6/\pi^2$, which in particular proves our result.

7.7. ★ Some of the most beautiful formulæ in the world IV

Hold on to your seats, for you're about to be taken on another journey through a beautiful world of mathematical formulas! In this section we derive many formulas found in Euler's wonderful book *Introduction to analysis of the infinite* [**65**]; his second book [**66**] is also great. We also give our tenth proof of Euler's formula for $\pi^2/6$ and our third proof of Gregory-Leibniz-Madhava's formula for $\pi/4$.

7.7.1. Bernoulli numbers and evaluating sums/products. We start our onslaught of beautiful formulæ with a formula for $\zeta(2k) = \sum_{n=1}^{\infty} \frac{1}{n^{2k}}$ in terms of Bernoulli numbers; this complements the formulæ in Section 5.3 when we didn't know about Bernoulli numbers. To find such a formula, we begin with the partial fraction expansion of the cotangent from Section 7.4:

$$\pi z \cot \pi z = 1 + 2z^2 \sum_{n=1}^{\infty} \frac{1}{z^2 - n^2} = 1 - 2\sum_{n=1}^{\infty} \frac{z^2}{n^2 - z^2}.$$

Next, we apply Cauchy's double series theorem to this sum. Let $z \in \mathbb{C}$ be near 0 and observe that

$$\frac{z^2}{n^2 - z^2} = \frac{z^2/n^2}{1 - z^2/n^2} = \sum_{k=1}^{\infty} \left(\frac{z^2}{n^2}\right)^k,$$

where we used the geometric series formula $\sum_{k=1}^{\infty} r^k = \frac{r}{1-r}$ for $|r| < 1$. Therefore,

$$\pi z \cot \pi z = 1 - 2\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \left(\frac{z^2}{n^2}\right)^k.$$

Since

$$\sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\left|\frac{z^2}{n^2}\right|^k = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\Big(\frac{|z|^2}{n^2}\Big)^k = \frac{1}{2}\Big(1-\pi\cot\pi|z|\Big) < \infty,$$

by Cauchy's double series theorem, we have

$$\pi z\cot\pi z = 1 - 2\sum_{k=1}^{\infty}\sum_{n=1}^{\infty}\Big(\frac{z^2}{n^2}\Big)^k = 1 - 2\sum_{k=1}^{\infty}\bigg(\sum_{n=1}^{\infty}\frac{1}{n^{2k}}\bigg)z^{2k}. \tag{7.35}$$

On the other hand, we recall from Section 6.8 that $z\cot z = \sum_{k=0}^{\infty}(-1)^k\frac{2^{2k}\,B_{2k}}{(2k)!}\,z^{2k}$ (for $|z|$ small), where the B_{2k}'s are the Bernoulli numbers. Replacing z with πz, we get

$$\pi z\cot\pi z = 1 + \sum_{k=1}^{\infty}(-1)^k\frac{2^{2k}\,B_{2k}}{(2k)!}\,\pi^{2k}\,z^{2k}.$$

Comparing this equation with (7.35) and using the identity theorem, we see that

$$-2\sum_{n=1}^{\infty}\frac{1}{n^{2k}} = (-1)^k\frac{2^{2k}\,B_{2k}}{(2k)!}\,\pi^{2k}, \qquad k=1,2,3,\ldots.$$

Rewriting this slightly, we obtain Euler's famous result: For $k=1,2,3,\ldots$,

$$\boxed{\sum_{n=1}^{\infty}\frac{1}{n^{2k}} = (-1)^{k-1}\frac{(2\pi)^{2k}\,B_{2k}}{2(2k)!}\quad;\quad\text{that is, }\ \zeta(2k) = (-1)^{k-1}\frac{(2\pi)^{2k}\,B_{2k}}{2(2k)!}.} \tag{7.36}$$

Using the known values of the Bernoulli numbers found in Section 6.8, setting $k=1,2,3$, we get, in particular, our eleventh proof of Euler's formula for $\pi^2/6$:

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty}\frac{1}{n^2}\quad\textbf{(Euler's sum, Proof XI)}\quad,\quad\frac{\pi^4}{90} = \sum_{n=1}^{\infty}\frac{1}{n^4}\quad,\quad\frac{\pi^6}{945} = \sum_{n=1}^{\infty}\frac{1}{n^6}.$$

Using (7.36), we can derive many other pretty formulas. First, in Theorem 7.14 we proved that

$$\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^z} = (1-2^{1-z})\zeta(z), \quad z>1.$$

In particular, setting $z=2k$, we find that for $k=1,2,3,\ldots$,

$$\boxed{\eta(2k) = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^{2k}} = (-1)^{k-1}\big(1-2^{1-2k}\big)\frac{(2\pi)^{2k}\,B_{2k}}{2(2k)!};} \tag{7.37}$$

what formulas do you get when you set $k=1,2$? Second, recall from Theorem 7.12 that

$$\sum_{n=1}^{\infty}\frac{1}{n^z} = \prod\frac{p^z}{p^z-1} = \frac{2^z}{2^z-1}\cdot\frac{3^z}{3^z-1}\cdot\frac{5^z}{5^z-1}\cdot\frac{7^z}{7^z-1}\cdots \tag{7.38}$$

where the product is over all primes. In particular, setting $z=2$, we get

$$\boxed{\frac{\pi^2}{6} = \frac{2^2}{2^2-1}\cdot\frac{3^2}{3^2-1}\cdot\frac{5^2}{5^2-1}\cdot\frac{7^2}{7^2-1}\cdot\frac{11^2}{11^2-1}\cdots} \tag{7.39}$$

and setting $z = 4$, we get

$$\boxed{\frac{\pi^4}{90} = \frac{2^4}{2^4-1} \cdot \frac{3^4}{3^4-1} \cdot \frac{5^4}{5^4-1} \cdot \frac{7^4}{7^4-1} \cdot \frac{11^4}{11^4-1} \cdots .}$$

Dividing these two formulas and using that

$$\frac{\dfrac{n^4}{n^4-1}}{\dfrac{n^2}{n^2-1}} = n^2 \cdot \frac{n^2-1}{n^4-1} = n^2 \cdot \frac{n^2-1}{(n^2-1)(n^2+1)} = \frac{n^2}{n^2+1},$$

we obtain

$$\boxed{\frac{\pi^2}{15} = \frac{2^2}{2^2+1} \cdot \frac{3^2}{3^2+1} \cdot \frac{5^2}{5^2+1} \cdot \frac{7^2}{7^2+1} \cdot \frac{11^2}{11^2+1} \cdots .} \tag{7.40}$$

Third, recall from Theorem 7.13 that

$$\frac{1}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^z},$$

where $\mu(n)$ is the Möbius function. In particular, setting $z = 2$, we find that

$$\boxed{\frac{6}{\pi^2} = 1 - \frac{1}{2^2} - \frac{1}{3^2} - \frac{1}{5^2} + \frac{1}{6^2} - \frac{1}{7^2} + \frac{1}{10^2} - \frac{1}{11^2} + \cdots ;}$$

what formula do you get when you set $z = 4$?

7.7.2. Euler numbers and evaluating sums.

7.7.2. Euler numbers and evaluating sums. We now derive a formula for the *alternating* sum of the odd natural numbers to odd powers:

$$1 - \frac{1}{3^{2k+1}} + \frac{1}{5^{2k+1}} - \frac{1}{7^{2k+1}} + \frac{1}{9^{2k+1}} - + \cdots , \qquad k = 0, 1, 2, 3, \ldots .$$

First try: To this end, let $|z| < 1$ and recall from Section 7.4 that

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \frac{1}{1^2 - z^2} - \frac{3}{3^2 - z^2} + \frac{5}{5^2 - z^2} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{(2n+1)}{(2n+1)^2 - z^2}. \tag{7.41}$$

Expanding as a geometric series, observe that

$$\frac{(2n+1)}{(2n+1)^2 - z^2} = \frac{1}{(2n+1)} \cdot \frac{1}{1 - \frac{z^2}{(2n+1)^2}} = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2n+1)^{2k+1}}. \tag{7.42}$$

Thus,

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-1)^n \frac{z^{2k}}{(2n+1)^{2k+1}}. \tag{7.43}$$

Just as we did in proving (7.35), we shall try to use Cauchy's double series theorem on this sum ... however, observe that

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \left| (-1)^n \frac{z^{2k}}{(2n+1)^{2k+1}} \right| = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{|z|^{2k}}{(2n+1)^{2k+1}} = \sum_{n=0}^{\infty} \frac{(2n+1)}{(2n+1)^2 - |z|^2},$$

which diverges (because this series behaves like $\sum \frac{1}{2n+1} = \infty$)! Therefore, we cannot apply Cauchy's double series theorem.

Second try: Let us start fresh from scratch. This time, we break up (7.41) into sums over n even and n odd (just consider the sums with n replaced by $2n$ and also by $2n+1$):

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \sum_{n=0}^{\infty}\left(\frac{(4n+1)}{(4n+1)^2 - z^2} - \frac{(4n+3)}{(4n+3)^2 - z^2}\right).$$

Let $|z| < 1$. Then writing $\frac{(4n+1)}{(4n+1)^2-z^2}$ and $\frac{(4n+3)}{(4n+3)^2-z^2}$ as geometric series (just as we did in (7.42)) we see that

$$\begin{aligned}\frac{\pi}{4\cos\frac{\pi z}{2}} &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\left(\frac{z^{2k}}{(4n+1)^{2k+1}} - \frac{z^{2k}}{(4n+3)^{2k+1}}\right)\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\left(\frac{1}{(4n+1)^{2k+1}} - \frac{1}{(4n+3)^{2k+1}}\right)z^{2k}. \qquad (7.44)\end{aligned}$$

We can now use Cauchy's double series theorem on this sum because

$$\begin{aligned}&\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\left|\left(\frac{1}{(4n+1)^{2k+1}} - \frac{1}{(4n+3)^{2k+1}}\right)z^{2k}\right|\\ &\qquad = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\left(\frac{1}{(4n+1)^{2k+1}} - \frac{1}{(4n+3)^{2k+1}}\right)|z|^{2k} = \frac{\pi}{4\cos\frac{\pi|z|}{2}} < \infty,\end{aligned}$$

where we used (7.44) with z replaced by $|z|$. Thus, by Cauchy's double series theorem, we have

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\left(\frac{1}{(4n+1)^{2k+1}} - \frac{1}{(4n+3)^{2k+1}}\right)z^{2k}.$$

Combining the terms in parentheses, we get

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\left(\frac{(-1)^n}{(2n+1)^{2k+1}}\right)z^{2k}; \qquad (7.45)$$

thus, we could in fact interchange orders in (7.43), but to justify it with complete mathematical rigor, we needed a little bit of mathematical gymnastics.

Now recall from Section 6.8 that

$$\frac{1}{\cos z} = \sec z = \sum_{k=0}^{\infty}(-1)^k\frac{E_{2k}}{(2k)!}\,z^{2k},$$

where the E_{2k}'s are the Euler numbers. Replacing z with $\pi z/2$ and multiplying by $\pi/4$, we get

$$\frac{\pi}{4\cos\frac{\pi z}{2}} = \frac{\pi}{4}\sum_{k=0}^{\infty}(-1)^k\frac{E_{2k}}{(2k)!}\left(\frac{\pi}{2}\right)^{2k} z^{2k}.$$

Comparing this equation with (7.45) and using the identity theorem, we conclude that for $k = 0, 1, 2, 3, \ldots$,

$$\boxed{\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^{2k+1}} = (-1)^k\frac{E_{2k}}{2(2k)!}\left(\frac{\pi}{2}\right)^{2k+1}.} \qquad (7.46)$$

In particular, setting $k = 0$ (and recalling that $E_0 = 1$) we get our third proof of Gregory-Leibniz-Madhava's formula:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots, \quad (\textbf{Gregory-Leibniz-Madhava, Proof III}).$$

What pretty formulas do you get when you set $k = 1, 2$? (Here, you need the Euler numbers calculated in Section 6.8.) We can derive many other pretty formulas from (7.46). To start this onslaught, we first state an "odd version" of Theorem 7.12:

Theorem 7.15. *For any $z \in \mathbb{C}$ with $\operatorname{Re} z > 1$, we have*

$$\boxed{\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^z} = \frac{3^z}{3^z+1} \cdot \frac{5^z}{5^z-1} \cdot \frac{7^z}{7^z+1} \cdot \frac{11^z}{11^z+1} \cdot \frac{13^z}{13^z-1} \cdots,}$$

where the product is over odd primes (all primes except 2) and where the $\pm$ signs in the denominators depends on whether the prime is of the form $4k+3$ ($+$ sign) or $4k+1$ ($-$ sign), where $k = 0, 1, 2, \ldots$.

Since the proof of this theorem is similar to that of Theorem 7.12, we shall leave the proof of this theorem to the interested reader; see Problem 5. In particular, setting $z = 1$, we get

$$\boxed{\frac{\pi}{4} = \frac{3}{4} \cdot \frac{5}{4} \cdot \frac{7}{8} \cdot \frac{11}{12} \cdot \frac{13}{12} \cdot \frac{17}{16} \cdot \frac{19}{20} \cdot \frac{23}{24} \cdots.} \tag{7.47}$$

The numerators of the fractions on the right are the odd prime numbers and the denominators are even numbers divisible by four and differing from the numerators by one. In (7.39), we found that

$$\frac{\pi^2}{6} = \frac{2^2}{2^2-1} \cdot \frac{3^2}{3^2-1} \cdot \frac{5^2}{5^2-1} \cdots = \frac{4}{3} \cdot \frac{3 \cdot 3}{2 \cdot 4} \cdot \frac{5 \cdot 5}{4 \cdot 6} \cdot \frac{7 \cdot 7}{6 \cdot 8} \cdot \frac{11 \cdot 11}{10 \cdot 12} \cdot \frac{13 \cdot 13}{12 \cdot 14} \cdots.$$

Dividing this expression by (7.47), and cancelling like terms, we obtain

$$\frac{4\pi}{6} = \frac{\pi^2/6}{\pi/4} = \frac{4}{3} \cdot \frac{3}{2} \cdot \frac{5}{6} \cdot \frac{7}{6} \cdot \frac{11}{10} \cdot \frac{13}{14} \cdot \frac{17}{18} \cdots.$$

Multiplying both sides by 3/4, we get another one of Euler's famous formulas:

$$\boxed{\frac{\pi}{2} = \frac{3}{2} \cdot \frac{5}{6} \cdot \frac{7}{6} \cdot \frac{11}{10} \cdot \frac{13}{14} \cdot \frac{17}{18} \cdot \frac{19}{18} \cdot \frac{23}{22} \cdots.} \tag{7.48}$$

The numerators of the fractions are the odd prime numbers and the denominators are even numbers not divisible by four and differing from the numerators by one. (7.47) and (7.48) are two of my favorite infinite product expansions for π.

7.7.3. Benoit Cloitre's e and π in a mirror. In this section we prove a unbelievable fact connecting e and π that is due to Benoit Cloitre [**52**], [**74**], [**205**]. Define sequences $\{a_n\}$ and $\{b_n\}$ by $a_1 = b_1 = 0$, $a_2 = b_2 = 1$, and the rest as the following "mirror images":

$$a_{n+2} = a_{n+1} + \frac{1}{n} a_n$$
$$b_{n+2} = \frac{1}{n} b_{n+1} + b_n.$$

We shall prove that

$$\boxed{e = \lim_{n\to\infty} \frac{n}{a_n} \qquad , \qquad \frac{\pi}{2} = \lim_{n\to\infty} \frac{n}{b_n^2}.} \tag{7.49}$$

The sequences $\{a_n\}$ and $\{b_n\}$ look so similar and so do $\{\frac{n}{a_n}\}$ and $\{\frac{n}{b_n^2}\}$, yet they generate very different numbers. Seeing such a connection between e and π, which *a priori* are very different, makes you wonder if there isn't someone behind this "coincidence."

To prove the formula for e, let us define a sequence $\{s_n\}$ by $s_n = a_n/n$. Then $s_1 = a_1/1 = 0$ and $s_2 = a_2/2 = 1/2$. Observe that for $n \geq 2$, we have

$$\begin{aligned} s_{n+1} - s_n = \frac{a_{n+1}}{n+1} - \frac{a_n}{n} &= \frac{1}{n+1}\Big(a_{n+1} - \frac{n+1}{n} a_n\Big) \\ &= \frac{1}{n+1}\Big(a_n + \frac{1}{n-1} a_{n-1} - \Big(1 + \frac{1}{n}\Big) a_n\Big) \\ &= \frac{1}{n+1}\Big(\frac{1}{n-1} a_{n-1} - \frac{a_n}{n}\Big) \\ &= \frac{-1}{n+1}(s_n - s_{n-1}). \end{aligned}$$

Using induction we see that

$$\begin{aligned} s_{n+1} - s_n = \frac{-1}{n+1}(s_n - s_{n-1}) &= \frac{-1}{n+1} \cdot \frac{-1}{n}(s_{n-1} - s_{n-2}) \\ &= \frac{-1}{n+1} \cdot \frac{-1}{n} \cdot \frac{-1}{n-1}(s_{n-2} - s_{n-3}) = \cdots \text{ etc.} \\ &= \frac{-1}{n+1} \cdot \frac{-1}{n} \cdot \frac{-1}{n-1} \cdots \frac{-1}{3}(s_2 - s_1) \\ &= \frac{-1}{n+1} \cdot \frac{-1}{n} \cdot \frac{-1}{n-1} \cdots \frac{-1}{3} \cdot \frac{1}{2} = \frac{(-1)^{n-3}}{(n+1)!} = \frac{(-1)^{n+1}}{(n+1)!}. \end{aligned}$$

Thus, writing as a telescoping sum, we obtain

$$s_n = s_1 + \sum_{k=2}^{n} (s_k - s_{k-1}) = 0 + \sum_{k=2}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

which is exactly the n-th partial sum for the series expansion of e^{-1}. It follows that $s_n \to e^{-1}$ and so,

$$\lim_{n\to\infty} \frac{n}{a_n} = \lim_{n\to\infty} \frac{1}{s_n} = \frac{1}{e^{-1}} = e,$$

as we claimed. The limit for π in (7.49) will be left to you (see Problem 2).

Exercises 7.7.

1. In this problem we derive other neat formulas:
   (1) Dividing (7.40) by $\pi^2/6$, prove that

$$\boxed{\frac{5}{2} = \frac{2^2+1}{2^2-1} \cdot \frac{3^2+1}{3^2-1} \cdot \frac{5^2+1}{5^2-1} \cdot \frac{7^2+1}{7^2-1} \cdot \frac{11^2+1}{11^2-1} \cdots,}$$

   quite a neat expression for 2.5.
   (2) Dividing (7.48) by (7.47), prove that

$$\boxed{2 = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \frac{8}{9} \cdot \frac{10}{9} \cdot \frac{12}{11} \cdots,}$$

quite a neat expression for 2. The fractions on the right are formed as follows: Given an odd prime $3, 5, 7, \ldots$, we take the pair of even numbers immediately above and below the prime, divide them by two, then put the resulting even number as the numerator and the odd number as the denominator.

2. In this problem, we prove the limit for π in (7.49).
   (i) Define $t_n = b_{n+1}/b_n$ for $n = 2, 3, 4, \ldots$. Prove that (for $n = 2, 3, 4, \ldots$), $t_{n+1} = 1/n + 1/t_n$ and then,
   $$t_n = \begin{cases} 1 & n \text{ even} \\ \frac{n}{n-1} & n \text{ odd.} \end{cases}$$
   (ii) Prove that $b_n^2 = t_2^2 \cdot t_3^2 \cdot t_4^2 \cdots t_{n-1}^2$, then using Wallis' formula, derive the limit for π in (7.49).
3. From Problem 7 in Exercises 7.6, prove that
   $$\frac{2(2n)!\,(1-2^{2n})}{(2\pi)^{2n}\,(1-2^{1-2n})} < |B_{2n}| < \frac{2(2n)!}{(2\pi)^{2n}\,(1-2^{1-2n})}.$$
   This estimate shows that the Bernoulli numbers grow incredibly fast as $n \to \infty$.
4. (**Radius of convergence**) In this problem we (finally) determine the radii of convergence of
   $$z \cot z = \sum_{n=0}^{\infty} (-1)^n \frac{2^{2n}\, B_{2n}}{(2n)!}\, z^{2n} \quad , \quad \tan z = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{2^{2n}(2^{2n}-1)\, B_{2n}}{(2n)!}\, z^{2n-1} .$$
   (a) Let $a_{2n} = (-1)^n \frac{2^{2n}\, B_{2n}}{(2n)!}$. Prove that
   $$\lim_{n\to\infty} |a_{2n}|^{1/2n} = \lim_{n\to\infty} \frac{1}{\pi} \cdot 2^{1/2n} \cdot \zeta(2n)^{1/2n} = \frac{1}{\pi}.$$
   Conclude that the radius of convergence of $z \cot z$ is π.
   (b) Using a similar argument, show that the radius of convergence of $\tan z$ is $\pi/2$.
5. In this problem, we prove Theorem 7.15
   (i) Let us call an odd number "type I" if it is of the form $4k+1$ for some $k = 0, 1, \ldots$ and "type II" if it is of the form $4k+3$ for some $k = 0, 1, \ldots$. Prove that every odd number is either of type I or type II.
   (ii) Prove that type I × type I = type I, type I × type II = type II, and type II × type II = type I.
   (iii) Let $a, b, \ldots, c \in \mathbb{N}$ be odd. Prove that if there is an *odd* number of type II integers amongst $a, b, \ldots, c$, then $a \cdot b \cdots c$ is of type II, otherwise, $a \cdot b \cdots c$ is type I.
   (iv) Show that
   $$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^z} = \sum_{n=0}^{\infty} \frac{1}{(4n+1)^z} - \sum_{n=0}^{\infty} \frac{1}{(4n+3)^z},$$
   a sum of type I and type II natural numbers!
   (v) Let $z \in \mathbb{C}$ with $\operatorname{Re} z \geq r > 1$, let $1 < N \in \mathbb{N}$, and let $3 < 5 < \cdots < m < 2N+1$ be the odd prime numbers less than $2N+1$. In a similar manner as in the proof of Theorem 7.12, show that
   $$\Big| \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n+1)^z} - \frac{3^z}{3^z+1} \cdot \frac{5^z}{5^z-1} \cdot \frac{7^z}{7^z+1} \cdot \frac{11^z}{11^z-1} \cdots \frac{m^z}{m^z \pm 1} \Big|$$
   $$\leq \sum_{n=N}^{\infty} \Big| \frac{1}{(2n+1)^z} \Big| \leq \sum_{n=N}^{\infty} \frac{1}{(2n+1)^r},$$
   where the + signs in the product are for type II odd primes and the − signs for type I odd primes. Now finish the proof of Theorem 7.15.

CHAPTER 8

Infinite continued fractions

From time immemorial, the infinite has stirred men's emotions more than any other question. Hardly any other idea has stimulated the mind so fruitfully ... In a certain sense, mathematical analysis is a symphony of the infinite.
David Hilbert (1862-1943) "On the infinite" [**24**].

We dabbed a little into the theory of continued fractions (that is, fractions that continue on and on and on ...) way back in the exercises of Section 3.4. In this chapter we concentrate on this fascinating subject. We start in Section 8.1 by showing that such fractions occur very naturally in long division and we give their basic definitions. In Section 5.3, we prove some pretty dramatic formulas (this is one reason continued fractions are so fascinating, at least to me). For example, we'll show that $4/\pi$ and π can be written as the continued fractions:

$$\boxed{\frac{4}{\pi} = 1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ddots}}}} \quad , \quad \pi = 3 + \cfrac{1^2}{6 + \cfrac{3^2}{6 + \cfrac{5^2}{6 + \cfrac{7^2}{6 + \ddots}}}}.}$$

The continued fraction on the left is due to Lord Brouncker (and is the first continued fraction ever recorded) and the one on the right is due to Euler. If you think these π formulas are cool, we'll derive the following formulas for e as well:

$$\boxed{e = 2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \cfrac{5}{5 + \ddots}}}} = 1 + \cfrac{1}{0 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \ddots}}}}}}}.}$$

We'll prove the formula on the left in Section 7.7, but you'll have to wait for the formula on the right until Section 8.7. In Section 8.3, we discuss elementary properties of continued fractions. In this section we also discuss how a Greek mathematician, Diophantus of Alexandrea (≈ 200–284 A.D.), can help you if you're stranded on an island with guys you can't trust and a monkey with a healthy appetite! In Section 8.4 we study the convergence properties of continued fractions.

Recall from our discussion on the amazing number π and its computations from ancient times (see Section 4.10) that throughout the years, the following approximation to π came up: 3, 22/7, 333/106, and 355/113. Did you ever wonder why these particular numbers occur? Also, did you ever wonder why our calendar is constructed the way it is (e.g. why leap years occur)? Finally, did you ever wonder why a piano has twelve keys (within an octave)? In Sections 8.5 and 8.6 you'll find out that these mysteries have to do with continued fractions! In Section 8.8 we study special types of continued fractions having to do with square roots and in Section 8.9 we learn why Archimedes needed around 8×10^{206544} cattle in order to "have abundant of knowledge in this science [mathematics]"!

In the very last section, Section 8.10, we look at continued fractions and transcendental numbers.

Chapter 8 objectives: The student will be able to ...

- define a continued fraction, state the Wallis-Euler and fundamental recurrence relations, and apply the continued fraction convergence theorem (Theorem 8.14).
- compute the canonical continued fraction of a given real number.
- understand the relationship between convergents of a simple continued fraction and best approximations, and the relationship between periodic simple continued fractions and quadratic irrationals.
- solve simple diophantine equations (of linear and Pell type).

8.1. Introduction to continued fractions

In this section we introduce the basics of continued fractions and see how they arise out of high school division and also from solving equations.

8.1.1. Continued fractions arise when dividing.

A common way continued fractions arise is through "repeated divisions".

Example 8.1. Take for instance, high school division of 68 into 157: $\frac{157}{68} = 2 + \frac{21}{68}$. Inverting the fraction $\frac{21}{68}$, we can write $\frac{157}{68}$ as

$$\frac{157}{68} = 2 + \cfrac{1}{\cfrac{68}{21}}.$$

Since we can further divide $\frac{68}{21} = 3 + \frac{5}{21} = 3 + \frac{1}{21/5}$, we can write $\frac{157}{68}$ in the somewhat fancy way

$$\frac{157}{68} = 2 + \cfrac{1}{3 + \cfrac{1}{\cfrac{21}{5}}}.$$

Since $\frac{21}{5} = 4 + \frac{1}{5}$, we can write

$$\frac{157}{68} = 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5}}}. \tag{8.1}$$

Since 5 is now a whole number, our repeated division process stops.

The expression on the right in (8.1) is called a **finite simple continued fraction**. There are many ways to denote the right-hand side, but we shall stick with the following two:

$$\langle 2; 3, 4, 5\rangle \quad \text{or} \quad 2 + \frac{1}{3+}\,\frac{1}{4+}\,\frac{1}{5} \qquad \text{represent} \quad 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5}}}.$$

Thus, continued fractions (that is, fractions that "continue on") arise naturally out of writing rational numbers in a somewhat fancy way by repeated divisions. Of course, 157 and 68 were not special, by repeated divisions one can take *any* two integers a and b with $b \neq 0$ and write a/b as a finite simple continued fraction; see Problem 2. In Section 8.4, we shall prove that any real number, not necessarily rational, can be expressed as a simple (possibly infinite) continued fraction.

8.1.2. Continued fractions arise when solving equations. Continued fractions also arise naturally when trying to solve equations.

Example 8.2. Suppose we want to find the positive solution x to the equation $x^2 - x - 2 = 0$. (Notice that $x = 2$ is the only positive solution.) On the other hand, writing $x^2 - x - 2 = 0$ as $x^2 = x + 2$ and dividing by x, we get

$$x = 1 + \frac{2}{x}.$$

We can replace x in the denominator with $x = 1 + 2/x$ to get

$$x = 1 + \cfrac{2}{1 + \cfrac{2}{x}}.$$

Repeating this many times, we can write

$$x = 1 + \cfrac{2}{1 + \cfrac{2}{1 + \cfrac{2}{\ddots\, 1 + \cfrac{2}{1 + \cfrac{2}{x}}}}}.$$

Repeating this "to infinity" and using that $x = 2$, we write

$$\text{“} \quad 2 = 1 + \cfrac{2}{1 + \cfrac{2}{1 + \cfrac{2}{1 + \cfrac{2}{1 + \ddots}}}}. \quad \text{”}$$

Quite a remarkable formula for 2. Later, (see Problem 4 in Section 8.4) we shall see that *any* whole number can be written in such a way. The reason for the quotation marks is that we have not yet defined what the right-hand object means.

However, we shall define what this means in a moment, but before doing so, here's another neat example:

Example 8.3. Consider the slightly modified formula $x^2 - x - 1 = 0$. Then $\Phi = \frac{1+\sqrt{5}}{2}$, called the **golden ratio**, is the only positive solution. We can rewrite $\Phi^2 - \Phi - 1 = 0$ as $\Phi = 1 + \frac{1}{\Phi}$. Replacing Φ in the denominator with $\Phi = 1 + \frac{1}{\Phi}$, we get

$$\Phi = 1 + \cfrac{1}{1 + \cfrac{1}{\Phi}}.$$

Repeating this substitution process "to infinity", we can write

$$\text{“}\quad \Phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}}}\quad ,\text{”} \tag{8.2}$$

quite a beautiful expression (cf. Problem 6 in Exercises 3.4)! As a side remark, there are many false rumors concerning the golden ratio; see [**146**] for the rundown.

8.1.3. Basic definitions for continued fractions. A general finite continued fraction can be written as

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cfrac{\ddots}{a_{n-1} + \cfrac{b_n}{a_n}}}}} \tag{8.3}$$

where the a_k's and b_k's are real numbers. (Of course, we are implicitly assuming that these fractions are all well-defined, e.g. no divisions by zero are allowed. Also, when you simplify this big fraction by combining fractions, you need to go from the bottom up.) Notice that if $b_m = 0$ for some m, then

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cfrac{\ddots}{a_{n-1} + \cfrac{b_n}{a_n}}}}} = a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{\ddots}{a_{m-2} + \cfrac{b_{m-1}}{a_{m-1}}}}}\ , \tag{8.4}$$

since the $b_m = 0$ will zero out everything below it. The continued fraction is called **simple** if all the b_k's are 1 and the a_k's are integers with a_k positive for $k \geq 1$. Instead of writing the continued fraction as we did above, which takes up a lot of space, we shall shorten it to:

$$a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\cdots\,\frac{b_n}{+\,a_n}.$$

In the case all $b_n = 1$, we shorten the notation to

$$a_0 + \frac{1}{a_1+} \frac{1}{a_2+} \frac{1}{a_3+} \cdots \frac{1}{+\, a_n} = \langle a_0; a_1, a_2, a_3, \ldots, a_n \rangle.$$

If $a_0 = 0$, some authors write $\langle a_1, a_2, \ldots, a_n \rangle$ instead of $\langle 0; a_1, \ldots a_n \rangle$.

We now discuss infinite continued fractions. Let $\{a_n\}$, $n = 0, 1, 2, \ldots$, and $\{b_n\}$, $n = 1, 2, \ldots$, be sequences of real numbers and suppose that

$$c_n := a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_n}{+\, a_n}$$

is defined for all n. We call c_n the **n-th convergent** of the continued fraction. If the limit, $\lim c_n$, exists, then we say that the **infinite continued fraction**

$$(8.5) \qquad a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \ddots}}} \quad \text{or} \quad a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots$$

converges and we use either of these notations to denote the limiting value $\lim c_n$. In the case all $b_n = 1$, in place of (8.5) we use the notation

$$\langle a_0; a_1, a_2, a_3, \ldots \rangle := \lim_{n \to \infty} \langle a_0; a_1, a_2, a_3, \ldots, a_n \rangle,$$

provided that the right-hand side exists. In Section 8.4 we shall prove that any simple continued fraction converges; in particular, we'll prove that (8.2) does hold true:

$$\Phi = 1 + \frac{1}{1+} \frac{1}{1+} \frac{1}{1+} \cdots.$$

In the case when there is some b_m term that vanishes, then convergence of (8.5) is easy because (see (8.4)) for $n \geq m$, we have $c_n = c_{m-1}$. Hence, in this case

$$a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots = a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_{m-1}}{+\, a_{m-1}}$$

converges automatically (such a continued fraction is said to **terminate** or be **finite**). However, general convergence issues are not so straightforward. We shall deal with the subtleties of convergence in Section 8.4.

Exercises 8.1.

1. Expand the following fractions into finite simple continued fractions:

$$(a)\ \frac{7}{11}\ ,\quad (b)\ -\frac{11}{7}\ ,\quad (c)\ \frac{3}{13}\ ,\quad (d)\ \frac{13}{3}\ ,\quad (e)\ -\frac{42}{31}.$$

2. Prove that a real number can be written as a finite simple continued fraction if and only if it is rational. Suggestion: for the "if" statement, use the division algorithm (see Theorem 2.15): For $a, b \in \mathbb{Z}$ with $b > 0$, we have $a = qb + r$ where $q, r \in \mathbb{Z}$ with $0 \leq r < b$; if a, b are both nonnegative, then so is q.
3. Let $\xi = a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_n}{+\, a_n} \neq 0$. Prove that

$$\frac{1}{\xi} = \frac{1}{a_0+} \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_n}{+\, a_n}.$$

In particular, if $\xi = \langle a_0; a_1, \ldots, a_n \rangle \neq 0$, show that $\frac{1}{\xi} = \langle 0; a_0, a_1, a_2, \ldots, a_n \rangle$.

4. A useful technique to study continued fraction is the following artifice of writing a continued fraction within a continued fraction. For a continued fraction $\xi = a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_n}{+ a_n}$, if $m < n$, prove that
$$\xi = a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_m}{+ \eta}, \quad \text{where } \eta = \frac{b_{m+1}}{a_{m+1}+} \cdots \frac{b_n}{+ a_n}.$$

8.2. ★ Some of the most beautiful formulæ in the world V

Hold on to your seats, for you're about to be taken on another journey through the beautiful world of mathematical formulas!

8.2.1. Transformation of continued fractions. It will often be convenient to transform one continued fraction to another one. For example, let ρ_1, ρ_2, ρ_3 be nonzero real numbers and suppose that the finite fraction
$$\xi = a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3}}},$$
where the a_k's and b_k's are real numbers, is defined. Then multiplying the top and bottom of the fraction by ρ_1, we get
$$\xi = a_0 + \cfrac{\rho_1 b_1}{\rho_1 a_1 + \cfrac{\rho_1 b_2}{a_2 + \cfrac{b_3}{a_3}}}.$$
Multiplying the top and bottom of the fraction with $\rho_1 b_2$ as numerator by ρ_2 gives
$$\xi = a_0 + \cfrac{\rho_1 b_1}{\rho_1 a_1 + \cfrac{\rho_1 \rho_2 b_2}{\rho_2 a_2 + \cfrac{\rho_2 b_3}{a_3}}}.$$
Finally, multiplying the top and bottom of the fraction with $\rho_2 b_3$ as numerator by ρ_3 gives
$$\xi = a_0 + \cfrac{\rho_1 b_1}{\rho_1 a_1 + \cfrac{\rho_1 \rho_2 b_2}{\rho_2 a_2 + \cfrac{\rho_2 \rho_3 b_3}{\rho_3 a_3}}}.$$
In summary,
$$a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3} = a_0 + \frac{\rho_1 b_1}{\rho_1 a_1+} \frac{\rho_1 \rho_2 b_2}{\rho_2 a_2 \ +} \frac{\rho_2 \rho_3 b_3}{\rho_3 a_3}.$$
A simple induction argument proves the following.

Theorem 8.1 (**Transformation rules**). *For any real numbers $a_1, a_2, a_3, \ldots$, $b_1, b_2, b_3, \ldots$, and nonzero constants $\rho_1, \rho_2, \rho_3, \ldots$, we have*
$$a_0 + \frac{b_1}{a_1+} \frac{b_2}{a_2+} \frac{b_3}{a_3+} \cdots \frac{b_n}{+ a_n} = a_0 + \frac{\rho_1 b_1}{\rho_1 a_1+} \frac{\rho_1 \rho_2 b_2}{\rho_2 a_2 \ +} \frac{\rho_2 \rho_3 b_3}{\rho_3 a_3 \ +} \cdots \frac{\rho_{n-1} \rho_n b_n}{+ \ \rho_n a_n},$$

in the sense when the left-hand side is defined, so is the right-hand side and this equality holds. In particular, if the limit as $n \to \infty$ of the left-hand side exists, then the limit of the right-hand side also exists, and

$$a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\cdots\,\frac{b_n}{+\,a_n +}\,\cdots = a_0 + \frac{\rho_1 b_1}{\rho_1 a_1 +}\,\frac{\rho_1\rho_2 b_2}{\rho_2 a_2\ +}\,\cdots\,\frac{\rho_{n-1}\rho_n b_n}{+\quad \rho_n a_n\quad +}\,\cdots.$$

8.2.2. Two stupendous series and continued fractions identities. Let $\alpha_1, \alpha_2, \alpha_3, \ldots$ be any real numbers with $\alpha_k \neq 0$ and $\alpha_k \neq \alpha_{k-1}$ for all k. Observe that

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_2} = \frac{\alpha_2 - \alpha_1}{\alpha_1\alpha_2} = \frac{1}{\frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1}}.$$

Since

$$\frac{\alpha_1\alpha_2}{\alpha_2 - \alpha_1} = \frac{\alpha_1(\alpha_2 - \alpha_1) + \alpha_1^2}{\alpha_2 - \alpha_1} = \alpha_1 + \frac{\alpha_1^2}{\alpha_2 - \alpha_1},$$

we get

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_2} = \frac{1}{\alpha_1 + \frac{\alpha_1^2}{\alpha_2 - \alpha_1}}.$$

This little exercise suggests the following theorem.

Theorem 8.2. *If $\alpha_1, \alpha_2, \alpha_3, \ldots$ are nonzero real numbers with $\alpha_k \neq \alpha_{k-1}$ for all k, then for any $n \in \mathbb{N}$,*

$$\sum_{k=1}^{n} \frac{(-1)^{k-1}}{\alpha_k} = \cfrac{1}{\alpha_1 + \cfrac{\alpha_1^2}{\alpha_2 - \alpha_1 + \cfrac{\alpha_2^2}{\ddots\, \alpha_3 - \alpha_2 + \cfrac{\alpha_{n-1}^2}{\alpha_n - \alpha_{n-1}}}}}.$$

In particular, taking $n \to \infty$, we conclude that

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{\alpha_k} = \frac{1}{\alpha_1 +}\,\frac{\alpha_1^2}{\alpha_2 - \alpha_1 +}\,\frac{\alpha_2^2}{\alpha_3 - \alpha_2 +}\,\frac{\alpha_3^2}{\alpha_4 - \alpha_3 +}\,\cdots \tag{8.6}$$

as long as either (and hence both) side makes sense.

Proof. This theorem is certainly true for alternating sums with $n = 1$ terms. Assume it is true for sums with n terms; we shall prove it holds for sums with $n+1$ terms. Observe that we can write

$$\begin{aligned}
\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{\alpha_k} &= \frac{1}{\alpha_1} - \frac{1}{\alpha_2} + \cdots + \frac{(-1)^{n-1}}{\alpha_n} + \frac{(-1)^n}{\alpha_{n+1}} \\
&= \frac{1}{\alpha_1} - \frac{1}{\alpha_2} + \cdots + (-1)^{n-1}\left(\frac{1}{\alpha_n} - \frac{1}{\alpha_{n+1}}\right) \\
&= \frac{1}{\alpha_1} - \frac{1}{\alpha_2} + \cdots + (-1)^{n-1}\left(\frac{\alpha_{n+1} - \alpha_n}{\alpha_n\alpha_{n+1}}\right) \\
&= \frac{1}{\alpha_1} - \frac{1}{\alpha_2} + \cdots + (-1)^{n-1}\frac{1}{\frac{\alpha_n\alpha_{n+1}}{\alpha_{n+1} - \alpha_n}}.
\end{aligned}$$

This is now a sum of n terms. Thus, we can apply the induction hypothesis to conclude that

$$\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{\alpha_k} = \frac{1}{\alpha_1 +} \; \frac{\alpha_1^2}{\alpha_2 - \alpha_1 +} \; \cdots \; \frac{\alpha_{n-1}^2}{+ \frac{\alpha_n \alpha_{n+1}}{\alpha_{n+1} - \alpha_n} - \alpha_{n-1}}. \tag{8.7}$$

Since

$$\begin{aligned} \frac{\alpha_n \alpha_{n+1}}{\alpha_{n+1} - \alpha_n} - \alpha_{n-1} &= \frac{\alpha_n(\alpha_{n+1} - \alpha_n) + \alpha_n^2}{\alpha_{n+1} - \alpha_n} - \alpha_{n-1} \\ &= \alpha_n - \alpha_{n-1} + \frac{\alpha_n^2}{\alpha_{n+1} - \alpha_n}, \end{aligned}$$

putting this into (8.7) gives

$$\sum_{k=1}^{n+1} \frac{(-1)^{k-1}}{\alpha_k} = \frac{1}{\alpha_1 +} \; \frac{\alpha_1^2}{\alpha_2 - \alpha_1 +} \; \cdots \; \frac{\alpha_{n-1}^2}{+ \alpha_n - \alpha_{n-1} + \frac{\alpha_n^2}{\alpha_{n+1} - \alpha_n}}.$$

This proves our induction step and completes our proof. □

Example 8.4. Since we know that

$$\log 2 = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots ,$$

setting $\alpha_k = k$ in (8.6), we can write

$$\boxed{\log 2 = \frac{1}{1+} \; \frac{1^2}{1\,+} \; \frac{2^2}{1\,+} \; \frac{3^2}{1\,+} \; \cdots ,}$$

which we can also write as the equally beautiful expression

$$\boxed{\log 2 = \cfrac{1}{1 + \cfrac{1^2}{1 + \cfrac{2^2}{1 + \cfrac{3^2}{1 + \cfrac{4^2}{1 + \ddots}}}}}.}$$

See Problem 1 for a general formula for $\log(1 + x)$.

Here is another interesting identity. Let $\alpha_1, \alpha_2, \alpha_3, \ldots$ be real, nonzero, and never equal to 1. Then observe that

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_1\alpha_2} = \frac{\alpha_2 - 1}{\alpha_1\alpha_2} = \frac{1}{\frac{\alpha_1\alpha_2}{\alpha_2 - 1}}.$$

Since

$$\frac{\alpha_1\alpha_2}{\alpha_2 - 1} = \frac{\alpha_1(\alpha_2 - 1) + \alpha_1}{\alpha_2 - 1} = \alpha_1 + \frac{\alpha_1}{\alpha_2 - 1},$$

we get

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_1\alpha_2} = \frac{1}{\alpha_1 + \frac{\alpha_1}{\alpha_2 - 1}}.$$

We can continue by induction in much the same manner as we did in the proof of Theorem 8.2 to obtain the following result.

Theorem 8.3. *For any real sequence $\alpha_1, \alpha_2, \alpha_3, \ldots$ with $\alpha_k \neq 0, 1$, we have*

$$\sum_{k=1}^{n} \frac{(-1)^{k-1}}{\alpha_1 \cdots \alpha_k} = \cfrac{1}{\alpha_1 + \cfrac{\alpha_1}{\alpha_2 - 1 + \cfrac{\alpha_2}{\alpha_3 - 1 + \cfrac{\ddots}{\alpha_{n-1} + \cfrac{\alpha_{n-1}}{\alpha_n - 1}}}}}.$$

In particular, taking $n \to \infty$, we conclude that

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{\alpha_1 \cdots \alpha_k} = \frac{1}{\alpha_1 +} \, \frac{\alpha_1}{\alpha_2 - 1 +} \, \frac{\alpha_2}{\alpha_3 - 1 +} \, \cdots \, \frac{\alpha_{n-1}}{+\, \alpha_n - 1 +} \, \cdots, \tag{8.8}$$

as long as either (and hence both) side makes sense.

Theorems 8.2 and 8.3 turn series to continued fractions; in Problem 9 we do the same for infinite products.

8.2.3. Continued fractions for arctan and π.

We now use the identities just learned to derive some remarkable continued fractions.

Example 8.5. First, since

$$\frac{\pi}{4} = \frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots,$$

using the limit expression (8.6) in Theorem 8.2:

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_2} + \frac{1}{\alpha_3} - \frac{1}{\alpha_4} + \cdots = \frac{1}{\alpha_1 +} \, \frac{\alpha_1^2}{\alpha_2 - \alpha_1 +} \, \frac{\alpha_2^2}{\alpha_3 - \alpha_2 +} \, \frac{\alpha_3^2}{\alpha_4 - \alpha_3 +} \, \cdots,$$

we can write

$$\frac{\pi}{4} = \cfrac{1}{1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ddots}}}}}.$$

Inverting both sides (see Problem 3 in Exercises 8.1), we obtain the incredible expansion:

$$\boxed{\frac{4}{\pi} = 1 + \cfrac{1^2}{2 + \cfrac{3^2}{2 + \cfrac{5^2}{2 + \cfrac{7^2}{2 + \ddots}}}}.} \tag{8.9}$$

This continued fraction was the very first continued fraction ever recorded, and was written down without proof by Lord Brouncker (1620 – 1686), the first president of the Royal Society of London.

Actually, we can derive (8.9) from a related expansion of the arctangent function, which is so neat that we shall derive in two ways, using Theorem 8.2 then using Theorem 8.3.

Example 8.6. Recall that

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + \cdots .$$

Setting $\alpha_1 = \frac{1}{x}$, $\alpha_2 = \frac{3}{x^3}$, $\alpha_3 = \frac{5}{x^5}$, and in general, $\alpha_n = \frac{2n-1}{x^{2n-1}}$ into the formula (8.6) from Theorem 8.2, we get the somewhat complicated formula

$$\arctan x = \frac{1}{\frac{1}{x}+}\,\frac{\frac{1}{x^2}}{\frac{3}{x^2}-\frac{1}{x}+}\,\frac{\left(\frac{3}{x^3}\right)^2}{\frac{5}{x^5}-\frac{3}{x^3}+}\,\cdots\,\frac{\left(\frac{2n-3}{x^{2n-3}}\right)^2}{+\,\frac{2n-1}{x^{2n-1}}-\frac{2n-3}{x^{2n-3}}+}\,\cdots .$$

However, we can simplify this using the transformation rule from Theorem 8.1:

$$\frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\,\cdots\,\frac{b_n}{+\,a_n+}\,\cdots = \frac{\rho_1 b_1}{\rho_1 a_1+}\,\frac{\rho_1\rho_2 b_2}{\rho_2 a_2\ +}\,\cdots\,\frac{\rho_{n-1}\rho_n b_n}{+\ \ \rho_n a_n\ \ +}\,\cdots .$$

(Here we drop the a_0 term from that theorem.) Let $\rho_1 = x$, $\rho_2 = x^3$, …, and in general, $\rho_n = x^{2n-1}$. Then,

$$\frac{1}{\frac{1}{x}+}\,\frac{\frac{1}{x^2}}{\frac{3}{x^3}-\frac{1}{x}+}\,\frac{\left(\frac{3}{x^3}\right)^2}{\frac{5}{x^5}-\frac{3}{x^3}+}\,\frac{\left(\frac{5}{x^5}\right)^2}{\frac{7}{x^7}-\frac{5}{x^5}+}\,\cdots = \frac{x}{1+}\,\frac{x^2}{3-x^2+}\,\frac{3^2x^2}{5-3x^2+}\,\frac{5^2x^2}{7-5x^2+}\,\cdots .$$

Thus,

$$\arctan x = \frac{x}{1+}\,\frac{x^2}{3-x^2+}\,\frac{3^2x^2}{5-3x^2+}\,\frac{5^2x^2}{7-5x^2+}\,\cdots,$$

or in a fancier way:

$$\boxed{\arctan x = \cfrac{x}{1+\cfrac{x^2}{(3-x^2)+\cfrac{3^2x^2}{(5-3x^2)+\cfrac{5^2x^2}{(7-5x^2)+\ddots}}}}}. \tag{8.10}$$

In particular, setting $x = 1$ and inverting, we get Lord Brouncker's formula:

$$\frac{4}{\pi} = 1 + \cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{7^2}{2+\ddots}}}}.$$

Example 8.7. We can also derive (8.10) using Theorem 8.3: Using once again that

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + \cdots$$

and setting $\alpha_1 = \frac{1}{x}$, $\alpha_2 = \frac{3}{x^2}$, $\alpha_3 = \frac{5}{3x^2}$, $\alpha_4 = \frac{7}{5x^2}$, $\cdots$, $\alpha_n = \frac{2n-1}{(2n-3)x^2}$ for $n \geq 2$, into the limiting formula (8.8) from Theorem 8.3:

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_1\alpha_2} + \frac{1}{\alpha_1\alpha_2\alpha_3} - \cdots = \frac{1}{\alpha_1+}\,\frac{\alpha_1}{\alpha_2-1+}\,\frac{\alpha_2}{\alpha_3-1+}\,\cdots\,\frac{\alpha_n}{+\,\alpha_{n+1}-1+}\,\cdots$$

we obtain

$$\arctan x = \frac{1}{\frac{1}{x}+} \; \frac{\frac{1}{x}}{\frac{3}{x^2}-1+} \; \frac{\frac{3}{x^2}}{\frac{5}{3x^2}-1+} \; \cdots \; + \frac{\frac{2n-1}{(2n-3)x^2}}{\frac{2n+1}{(2n-1)x^2}-1+} \; \cdots.$$

From Theorem 8.1, we know that

$$\frac{b_1}{a_1+} \; \frac{b_2}{a_2+} \; \cdots \; \frac{b_n}{+a_n+} \; \cdots = \frac{\rho_1 b_1}{\rho_1 a_1+} \; \frac{\rho_1\rho_2 b_2}{\rho_2 a_2} + \; \cdots + \; \frac{\rho_{n-1}\rho_n b_n}{\rho_n a_n} + \; \cdots.$$

In particular, setting $\rho_1 = x$, $\rho_2 = x^2$, $\rho_3 = 3x^2$, $\rho_4 = 5x^2$, and in general, $\rho_n = (2n-3)x^2$ for $n \geq 1$ into the formula for $\arctan x$, we obtain (as you are invited to verify) the exact same expression (8.10)!

Example 8.8. We leave the next two beauts to you! Applying Theorem 8.2 and/or Theorem 8.3 to Euler's sum $\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$, in Problem 2 we ask you to derive the formula

$$\boxed{\frac{6}{\pi^2} = 0^2 + 1^2 - \cfrac{1^4}{1^2 + 2^2 - \cfrac{2^4}{2^2+3^2 - \cfrac{3^4}{3^2+4^2 - \cfrac{4^4}{4^2+5^2-\ddots}}}},} \tag{8.11}$$

which is, after inversion, the last formula on the front cover of this book.

Example 8.9. In Problem 9 we derive Euler's splendid formula [**47**, p. 89]:

$$\boxed{\frac{\pi}{2} = 1 + \cfrac{1}{1 + \cfrac{1\cdot 2}{1 + \cfrac{2\cdot 3}{1 + \cfrac{3\cdot 4}{1+\ddots}}}}.} \tag{8.12}$$

8.2.4. Another continued fraction for π. We now derive another remarkable formula for π, which is due to Euler (according to Castellanos [**47**, p. 89]; the proof we give is found in Lange's article [**129**]). Consider first the telescoping sum

$$\sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{1}{n}+\frac{1}{n+1}\right) = \left(\frac{1}{1}+\frac{1}{2}\right) - \left(\frac{1}{2}+\frac{1}{3}\right) + \left(\frac{1}{3}+\frac{1}{4}\right) - + \cdots = 1.$$

Since

$$\frac{\pi}{4} = \frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = 1 - \sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n+1},$$

multiplying this expression by 4 and using the previous expression, we can write

$$\begin{aligned}\pi = 4 - 4\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n+1} &= 3 + 1 - 4\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n+1}\\ &= 3 + \sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{1}{n}+\frac{1}{n+1}\right) - 4\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n+1}\\ &= 3 + \sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{1}{n}+\frac{1}{n+1}-\frac{4}{2n+1}\right)\\ &= 3 + 4\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n(2n+1)(2n+2)},\end{aligned}$$

where we combined fractions in going from the third to forth lines. We now apply the limiting formula (8.6) from Theorem 8.2 with $\alpha_n = 2n(2n+1)(2n+2)$. Observe that

$$\begin{aligned}\alpha_n - \alpha_{n-1} &= 2n(2n+1)(2n+2) - 2(n-1)(2n-1)(2n)\\ &= 4n\big[(2n+1)(n+1) - (n-1)(2n-1)\big]\\ &= 4n\big[2n^2+2n+n+1-(2n^2-n-2n+1)\big] = 4n(6n) = 24n^2.\end{aligned}$$

Now putting the α_n's into the formula

$$\frac{1}{\alpha_1}-\frac{1}{\alpha_2}+\frac{1}{\alpha_3}-\frac{1}{\alpha_4}+\cdots = \frac{1}{\alpha_1+}\,\frac{\alpha_1^2}{\alpha_2-\alpha_1+}\,\frac{\alpha_2^2}{\alpha_3-\alpha_2+}\,\frac{\alpha_3^2}{\alpha_4-\alpha_3+}\cdots,$$

we get

$$\begin{aligned}4\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{2n(2n+1)(2n+2)} &= 4\cdot\left(\frac{1}{2\cdot3\cdot4+}\,\frac{(2\cdot3\cdot4)^2}{24\cdot2^2+}\,\frac{(4\cdot5\cdot6)^2}{24\cdot3^2+}\cdots\right)\\ &= \frac{1}{2\cdot3+}\,\frac{(2\cdot3)^2\cdot4}{24\cdot2^2+}\,\frac{(4\cdot5\cdot6)^2}{24\cdot3^2+}\cdots.\end{aligned}$$

Hence,

$$\pi = 3+\frac{1}{6+}\,\frac{(2\cdot3)^2\cdot4}{24\cdot2^2+}\,\frac{(4\cdot5\cdot6)^2}{24\cdot3^2+}\cdots\frac{(2(n-1)(2n-1)(2n))^2}{24\cdot n^2+}\cdots,$$

which is beautiful, but we can make this even more beautiful using the transformation rule from Theorem 8.1:

$$\frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\cdots\frac{b_n}{a_n+}\cdots = \frac{\rho_1 b_1}{\rho_1 a_1+}\,\frac{\rho_1\rho_2 b_2}{\rho_2 a_2+}\cdots\frac{\rho_{n-1}\rho_n b_n}{\rho_n a_n+}\cdots.$$

Setting $\rho_1 = 1$ and $\rho_n = \frac{1}{4n^2}$ for $n \geq 2$, we see that for $n \geq 3$ we have

$$\frac{\rho_{n-1}\rho_n b_n}{\rho_n a_n} = \frac{\frac{1}{4(n-1)^2}\cdot\frac{1}{4n^2}\cdot(2(n-1)(2n-1)(2n))^2}{\frac{1}{4n^2}\cdot24\cdot n^2} = \frac{(2n-1)^2}{6};$$

the same formula holds for $n = 2$ as you can check. Thus,

$$\pi = 3+\frac{1^2}{6+}\,\frac{3^2}{6+}\,\frac{5^2}{6+}\,\frac{7^2}{6+}\cdots\frac{(2n-1)^2}{6+}\cdots$$

or in more elegant notation:

$$\boxed{\pi = 3 + \cfrac{1^2}{6 + \cfrac{3^2}{6 + \cfrac{5^2}{6 + \cfrac{7^2}{6 + \ddots}}}}.} \tag{8.13}$$

8.2.5. Continued fractions and e. For our final beautiful example, we shall compute a continued fraction expansion for e. We begin with

$$\frac{1}{e} = e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} = 1 - \frac{1}{1} + \frac{1}{1 \cdot 2} - \frac{1}{1 \cdot 2 \cdot 3} + \cdots ,$$

so

$$\frac{e-1}{e} = 1 - \frac{1}{e} = \frac{1}{1} - \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} - \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + \cdots .$$

Thus, setting $\alpha_k = k$ into the formula (8.8):

$$\frac{1}{\alpha_1} - \frac{1}{\alpha_1\alpha_2} + \frac{1}{\alpha_1\alpha_2\alpha_3} - \cdots = \frac{1}{\alpha_1 +} \; \frac{\alpha_1}{\alpha_2 - 1 +} \; \frac{\alpha_2}{\alpha_3 - 1 +} \; \cdots \; \frac{\alpha_{n-1}}{+ \alpha_n - 1 +} \; \cdots ,$$

we obtain

$$\frac{e-1}{e} = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{2}{2 + \cfrac{3}{3 + \ddots}}}} .$$

We can make this into an expression for e as follows: First, invert the expression and then subtract 1 from both sides to get

$$\frac{e}{e-1} = 1 + \cfrac{1}{1 + \cfrac{2}{2 + \cfrac{3}{3 + \ddots}}} , \quad \text{then} \quad \frac{1}{e-1} = \cfrac{1}{1 + \cfrac{2}{2 + \cfrac{3}{3 + \ddots}}} .$$

Second, invert again to obtain

$$e - 1 = 1 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \cfrac{5}{5 + \ddots}}}} .$$

Finally, adding 1 to both sides we get the incredibly beautiful expression

$$\boxed{e = 2 + \cfrac{2}{2 + \cfrac{3}{3 + \cfrac{4}{4 + \cfrac{5}{5 + \ddots}}}},} \tag{8.14}$$

or in shorthand:

$$\boxed{e = 2 + \frac{2}{2+}\,\frac{3}{3+}\,\frac{4}{4+}\,\frac{5}{5+}\cdots.}$$

In the exercises you will derive other amazing formulæ.

Exercises 8.2.

1. Recall that $\log(1+x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{n+1}}{n+1}$. Using this formula, the formula (8.6) derived from Theorem 8.2, and also the transformation rule, prove that fabulous formula

$$\boxed{\log(1+x) = \cfrac{x}{1 + \cfrac{1^2 x}{(2-1x) + \cfrac{2^2 x}{(3-2x) + \cfrac{3^2 x}{(4-3x) + \ddots}}}}.}$$

Plug in $x = 1$ to derive our beautiful formula for $\log 2$.

2. Using Euler's sum $\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$, give two proofs of the formula (8.11), one using Theorem 8.2 and the other using Theorem 8.3. The transformation rules will come in handy.

3. (i) For any real numbers $\{\alpha_k\}$, prove that for any n,

$$\sum_{k=0}^{n} \alpha_k x^k = \alpha_0 + \frac{\alpha_1 x}{1\;+}\,\frac{-\frac{\alpha_2}{\alpha_1}x}{1 + \frac{\alpha_2}{\alpha_1}x +}\,\frac{-\frac{\alpha_3}{\alpha_2}x}{1 + \frac{\alpha_3}{\alpha_2}x +}\cdots\frac{-\frac{\alpha_n}{\alpha_{n-1}}x}{+\,1 + \frac{\alpha_n}{\alpha_{n-1}}x}$$

provided, of course, that the right-hand side is defined, which we assume holds for every n.

(ii) Deduce that if the infinite series $\sum_{n=0}^{\infty} \alpha_n x^n$ converges, then

$$\sum_{n=0}^{\infty} \alpha_n x^n = \alpha_0 + \frac{\alpha_1 x}{1\;+}\,\frac{-\frac{\alpha_2}{\alpha_1}x}{1 + \frac{\alpha_2}{\alpha_1}x +}\,\frac{-\frac{\alpha_3}{\alpha_2}x}{1 + \frac{\alpha_3}{\alpha_2}x +}\cdots\frac{-\frac{\alpha_n}{\alpha_{n-1}}x}{+\,1 + \frac{\alpha_n}{\alpha_{n-1}}x +}\cdots.$$

Transforming the continued fraction on the right, prove that

$$\sum_{n=0}^{\infty} \alpha_n x^n = \alpha_0 + \frac{\alpha_1 x}{1\;+}\,\frac{-\alpha_2 x}{\alpha_1 + \alpha_2 x +}\,\frac{-\alpha_1\alpha_3 x}{\alpha_2 + \alpha_3 x +}\cdots\frac{-\alpha_{n-2}\alpha_n x}{+\,\alpha_{n-1} + \alpha_n x +}\cdots.$$

4. Writing $\arctan x = x(1 - \frac{y}{3} + \frac{y^2}{5} - \frac{y^3}{7} + \cdots)$ where $y = x^2$, and using the previous problem on $(1 - \frac{y}{3} + \frac{y^2}{5} - \frac{y^3}{7} + \cdots)$, derive the formula (8.10).

5. Let $x, y > 0$. Prove that

$$\boxed{\sum_{n=0}^{\infty} \frac{(-1)^n}{x+ny} = \frac{1}{x+}\,\frac{x^2}{y\,+}\,\frac{(x+y)^2}{y\quad+}\,\frac{(x+2y)^2}{y\quad+}\,\frac{(x+3y)^2}{y\quad+}\cdots.}$$

Suggestion: The formula (8.6) might help.

6. Recall the partial fraction expansion $\pi x \cot \pi x = 1 + 2x^2 \sum_{n=1}^{\infty} \frac{1}{x^2 - n^2}$.

(a) By breaking up $\frac{2x}{x^2-n^2}$ using partial fractions, prove that

$$\pi\cot\pi x = \frac{1}{x} - \frac{1}{1-x} + \frac{1}{1+x} - \frac{1}{2-x} + \frac{1}{2+x} - + \cdots .$$

(b) Derive the remarkable formula

$$\boxed{\pi\cot\pi x = \frac{1}{x+}\,\frac{x^2}{1-2x+}\,\frac{(1-x)^2}{2x\ \ +}\,\frac{(1+x)^2}{1-2x\ +}\,\frac{(2-x)^2}{2x\ \ +}\,\frac{(2+x)^2}{1-2x\ +}\cdots .}$$

Putting $x = 1/4$, give a new proof of Lord Brouncker's formula.

(c) Derive

$$\boxed{\frac{\tan\pi x}{\pi x} = 1 + \frac{x}{1-2x+}\,\frac{(1-x)^2}{2x\ \ +}\,\frac{(1+x)^2}{1-2x\ +}\,\frac{(2-x)^2}{2x\ \ +}\,\frac{(2+x)^2}{1-2x\ +}\cdots .}$$

7. Recall that $\frac{\pi}{\sin\pi x} = \frac{1}{x} + \sum_{n=1}^{\infty}\frac{2x}{n^2-x^2}$. From this, derive the beautiful expression

$$\boxed{\frac{\sin\pi x}{\pi x} = 1 - \frac{x}{1+}\,\frac{(1-x)^2}{2x\ \ +}\,\frac{(1+x)^2}{1-2x\ +}\,\frac{(2-x)^2}{2x\ \ +}\,\frac{(2+x)^2}{1-2x\ +}\cdots .}$$

Suggestion: First break up $\frac{2x}{n^2-x^2}$ and use an argument as you did for $\pi\cot\pi x$ to get a continued fraction expansion for $\frac{\pi}{\sin\pi x}$. From this, deduce the continued fraction expansion for $\sin\pi x/\pi x$.

8. From the expansion $\frac{\pi}{4\cos\frac{\pi x}{2}} = \sum_{n=0}^{\infty}(-1)^n\frac{(2n+1)}{(2n+1)^2-x^2}$ derive the beautiful expression

$$\boxed{\frac{\cos\frac{\pi x}{2}}{\frac{\pi}{2}} = x+1+\frac{(x+1)^2}{-2\cdot 1\ +}\,\frac{(x-1)^2}{-2\ \ +}\,\frac{(x-3)^2}{2\cdot 3\ \ +}\,\frac{(x+3)^2}{2\ \ +}\,\frac{(x+5)^2}{-2\cdot 5\ +}\,\frac{(x-5)^2}{-2\ \ +}\cdots .}$$

9. (Cf. [**114**]) In this problem we turn infinite products to continued fractions.

(a) Let $\alpha_1, \alpha_2, \alpha_3, \ldots$ be a sequence of real numbers with $\alpha_k \neq 0, -1$. Define sequences $b_1, b_2, b_3, \ldots$ and $a_0, a_1, a_2, \ldots$ by $b_1 = (1+\alpha_0)\alpha_1$, $a_0 = 1+\alpha_0$, $a_1 = 1$, and

$$b_n = -(1+\alpha_{n-1})\frac{\alpha_n}{\alpha_{n-1}} \quad , \quad \alpha_n = 1 - a_n \quad \text{for } n = 2, 3, 4, \ldots .$$

Prove (say by induction) that for any $n \in \mathbb{N}$,

$$\prod_{k=0}^{n}(1+\alpha_k) = a_0 + \frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\cdots\frac{b_n}{+\,a_n}.$$

Taking $n \to \infty$, get a formula between infinite products and fractions.

(b) Using that $\frac{\sin\pi x}{\pi x} = \prod_{n=1}^{\infty}\Big(1-\frac{x}{n}\Big) = (1-x)(1+x)\Big(1-\frac{x}{2}\Big)\Big(1+\frac{x}{2}\Big)\Big(1-\frac{x}{3}\Big)\Big(1+\frac{x}{3}\Big)\cdots$ and (a), derive the continued fraction expansion

$$\boxed{\frac{\sin\pi x}{\pi x} = 1 - \frac{x}{1+}\,\frac{1\cdot(1-x)}{x\ \ +}\,\frac{1\cdot(1+x)}{1-x\ \ +}\,\frac{2\cdot(2-x)}{x\ \ +}\,\frac{2\cdot(2+x)}{1-x\ \ +}\cdots .}$$

(c) Putting $x = 1/2$, prove (8.12). Putting $x = -1/2$, derive another continued fraction for $\pi/2$.

8.3. Recurrence relations, Diophantus' tomb, and shipwrecked sailors

In this section we define the Wallis-Euler recurrence relations, which generate sequences of numerators and denominators for convergents of continued fractions. Diophantine equations is the subject of finding integer solutions to polynomial equations. Continued fractions (through the special properties of the Wallis-Euler recurrence relations) turn out to play a very important role in this subject.

8.3.1. Convergents and recurrence relations. We shall call a continued fraction

$$a_0 + \frac{b_1}{a_1 +} \frac{b_2}{a_2 +} \frac{b_3}{a_3 +} \cdots \frac{b_n}{+ a_n +} \cdots \tag{8.15}$$

nonnegative if the a_n, b_n's are real numbers with $a_n > 0, b_n \geq 0$ for all $n \geq 1$ (a_0 can be any sign). We shall not spend a lot of time on continued fractions when the a_n's and b_n's in (8.15), for $n \geq 1$, are arbitrary real numbers; it is only for nonnegative infinite continued fractions that we develop their convergence properties in Section 8.4. However, we shall come across continued fractions where some of the a_n, b_n are negative — see for instance the beautiful expression (8.44) for $\cot x$ (and the following one for $\tan x$) in Section 8.7! We focus on continued fractions with $a_n, b_n > 0$ for $n \geq 1$ in order to avoid some possible contradictory statements. For instance, the convergents of the elementary example $\frac{1}{1+} \frac{-1}{1} + \frac{1}{1}$ have some weird properties. Let us form its convergents: $c_1 = 1$, which is OK, but

$$c_2 = \frac{1}{1+} \frac{-1}{1} = \cfrac{1}{1 + \cfrac{-1}{1}} = \frac{1}{1-1} = \frac{1}{0} = ???,$$

which is not OK.[1] However,

$$c_3 = \frac{1}{1+} \frac{-1}{1} \frac{1}{+1} = \cfrac{1}{1 + \cfrac{-1}{1 + \cfrac{1}{1}}} = \cfrac{1}{1 + \cfrac{-1}{2}} = \cfrac{1}{\cfrac{1}{2}} = 2,$$

which is OK again! To avoid such craziness, we shall focus on continued fractions with $a_n > 0$ for $n \geq 1$ and $b_n \geq 0$, but *we emphasize that much of what we do in this section and the next works in greater generality.*

Let $\{a_n\}_{n=0}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$ be sequences of real numbers with $a_n > 0, b_n \geq 0$ for all $n \geq 1$ (there is no restriction on a_0). The following sequences $\{p_n\}$, $\{q_n\}$ are central in the theory of continued fractions:

$$\boxed{\begin{array}{l} p_n = a_n p_{n-1} + b_n p_{n-2} \quad , \quad q_n = a_n q_{n-1} + b_n q_{n-2} \\ p_{-1} = 1 \ , \ p_0 = a_0 \quad , \quad q_{-1} = 0 \ , \ q_0 = 1. \end{array}} \tag{8.16}$$

We shall call these recurrence relations the **Wallis-Euler recurrence relations** ... you'll see why they're so central in a moment. In particular,

$$\boxed{\begin{array}{l} p_1 = a_1 p_0 + b_1 p_{-1} = a_1 a_0 + b_1 \\ q_1 = a_1 q_0 + b_1 q_{-1} = a_1. \end{array}} \tag{8.17}$$

We remark that $q_n > 0$ for $n = 0, 1, 2, 3, \ldots$. This is easily proved by induction: Certainly, $q_0 = 1, q_1 = a_1 > 0$ (recall that $a_n > 0$ for $n \geq 1$); thus assuming that $q_n > 0$ for $n = 0, \ldots, n-1$, we have (recall that $b_n \geq 0$),

$$q_n = a_n q_{n-1} + b_n q_{n-2} > 0 \cdot 0 + 0 = 0,$$

[1] Actually, in the continued fraction community, we always define $a/0 = \infty$ for $a \neq 0$ so we can get around this division by zero predicament by simply defining it away.

and our induction is complete. Observe that the zero-th convergent of the continued fraction (8.15) is $c_0 = a_0 = p_0/q_0$ and the first convergent is

$$c_1 = a_0 + \frac{b_1}{a_1} = \frac{a_1 a_0 + b_1}{a_1} = \frac{p_1}{q_1}.$$

The central property of the p_n, q_n's is the fact that $c_n = p_n/q_n$ for all n.

Theorem 8.4. *For any positive real number x, we have*

$$\boxed{a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\cdots\,\frac{b_n}{+\ x} = \frac{x p_{n-1} + b_n p_{n-2}}{x q_{n-1} + b_n q_{n-2}}, \qquad n = 1, 2, 3, \ldots.} \tag{8.18}$$

(Note that the denominator is > 0 because $q_n > 0$ for $n \geq 0$.) In particular, setting $x = a_n$ and using the definition of p_n, q_n, we have

$$c_n = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\cdots\,\frac{b_n}{+\ a_n} = \frac{p_n}{q_n}, \qquad n = 0, 1, 2, 3, \ldots.$$

Proof. We prove (8.18) by induction on the number of terms after a_0. The proof for one term after a_0 is easy: $a_0 + \frac{b_1}{x} = \frac{a_0 x + b_1}{x} = \frac{x p_0 + b_1 p_{-1}}{x q_0 + q_{-1}}$, since $p_0 = a_0$, $p_{-1} = 1$, $q_0 = 1$, and $q_{-1} = 0$. Assume that (8.18) holds when there are n terms after a_0; we shall prove it holds for fractions with $n + 1$ terms after a_0. To do so, we write (see Problem 4 in Exercises 8.1 for the general technique)

$$a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\cdots\,\frac{b_n}{+\ a_n +}\,\frac{b_{n+1}}{x} = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\cdots\,\frac{b_n}{+\ y},$$

where

$$y := a_n + \frac{b_{n+1}}{x} = \frac{x a_n + b_{n+1}}{x}.$$

Therefore, by our induction hypothesis, we have

$$\begin{aligned}
a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\cdots\,\frac{b_{n+1}}{+\ x} = \frac{y p_{n-1} + b_n p_{n-2}}{y q_{n-1} + b_n q_{n-2}} &= \frac{\left(\dfrac{x a_n + b_{n+1}}{x}\right) p_{n-1} + b_n p_{n-2}}{\left(\dfrac{x a_n + b_{n+1}}{x}\right) q_{n-1} + b_n q_{n-2}} \\
&= \frac{x a_n p_{n-1} + b_{n+1} p_{n-1} + x b_n p_{n-2}}{x a_n q_{n-1} + b_{n+1} q_{n-1} + x b_n q_{n-2}} \\
&= \frac{x(a_n p_{n-1} + b_n p_{n-2}) + b_{n+1} p_{n-1}}{x(a_n q_{n-1} + b_n q_{n-2}) + b_{n+1} q_{n-1}} \\
&= \frac{x p_n + b_{n+1} p_{n-1}}{x q_n + b_{n+1} q_{n-1}},
\end{aligned}$$

which completes our induction step and finishes our proof. □

In the next theorem, we give various useful identities that the p_n, q_n satisfy.

Theorem 8.5 (**Fundamental recurrence relations**). *For all $n \geq 1$, the following identities hold:*

$$\boxed{\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (-1)^{n-1} b_1 b_2 \cdots b_n \\ p_n q_{n-2} - p_{n-2} q_n &= (-1)^n a_n b_1 b_2 \cdots b_{n-1} \end{aligned}}$$

and (where the formula for $c_n - c_{n-2}$ is only valid for $n \geq 2$)

$$\boxed{c_n - c_{n-1} = \frac{(-1)^{n-1} b_1 b_2 \cdots b_n}{q_n\, q_{n-1}} \quad , \quad c_n - c_{n-2} = \frac{(-1)^n a_n b_1 b_2 \cdots b_{n-1}}{q_n\, q_{n-2}}.}$$

Proof. To prove that $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1} b_1 b_2 \cdots b_n$ for $n = 1, 2, \ldots$, we proceed by induction. For $n = 1$, the left-hand side is (see (8.17))

$$p_1 q_0 - p_0 q_1 = (a_1 a_0 + b_1) \cdot 1 - a_0 \cdot a_1 = b_1,$$

which is the right-hand side when $n = 1$. Assume our equality holds for n; we prove it holds for $n + 1$. By the Wallis-Euler recurrence relations, we have

$$\begin{aligned} p_{n+1} q_n - p_n q_{n+1} &= \big(a_{n+1} p_n + b_{n+1} p_{n-1}\big) q_n - p_n \big(a_{n+1} q_n + b_{n+1} q_{n-1}\big) \\ &= b_{n+1} p_{n-1} q_n - p_n b_{n+1} q_{n-1} \\ &= -b_{n+1} \big(p_n q_{n-1} - p_{n-1} q_n\big) \\ &= -b_{n+1} \cdot (-1)^{n-1} b_1 b_2 \cdots b_n = (-1)^n b_1 b_2 \cdots b_n b_{n+1}, \end{aligned}$$

which completes our induction step. To prove the second equality, we use the Wallis-Euler recurrence relations and the equality just proved:

$$\begin{aligned} p_n q_{n-2} - p_{n-2} q_n &= \big(a_n p_{n-1} + b_n p_{n-2}\big) q_{n-2} - p_{n-2} \big(a_n q_{n-1} + b_n q_{n-2}\big) \\ &= a_n p_{n-1} q_{n-2} - p_{n-2} a_n q_{n-1} \\ &= a_n \big(p_{n-1} q_{n-2} - p_{n-2} q_{n-1}\big) \\ &= a_n \cdot (-1)^{n-2} b_1 b_2 \cdots b_{n-1} = (-1)^n a_n b_1 b_2 \cdots b_{n-1}. \end{aligned}$$

Finally, the equations for $c_n - c_{n-1}$ and $c_n - c_{n-2}$ follow from dividing

$$\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (-1)^{n-1} b_1 b_2 \cdots b_n \\ p_n q_{n-2} - p_{n-2} q_n &= (-1)^{n-1} a_n b_1 b_2 \cdots b_{n-1} \end{aligned}$$

by $q_n\, q_{n-1}$ and $q_n\, q_{n-2}$, respectively. □

For simple continued fractions, the Wallis-Euler relations (8.16) and (8.17) and the fundamental recurrence relations take the following particularly simple forms:

Corollary 8.6 (**Simple fundamental recurrence relations**). *For simple continued fractions, for all $n \geq 1$, if*

$$\boxed{\begin{matrix} p_n = a_n p_{n-1} + p_{n-2} & , & q_n = a_n q_{n-1} + q_{n-2} \\ p_0 = a_0 \ , \ p_1 = a_0 a_1 + 1 & , & q_0 = 1 \ , \ q_1 = a_1, \end{matrix}}$$

then $c_n = p_n / q_n$ for all $n \geq 0$, and for any $x > 0$,

$$\boxed{\langle a_0; a_1, a_2, a_3, \ldots, a_n, x\rangle = \frac{x p_{n-1} + p_{n-2}}{x q_{n-1} + q_{n-2}}, \quad n = 1, 2, 3, \ldots.} \tag{8.19}$$

Moreover, for all $n \geq 1$, the following identities hold:

$$\boxed{\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (-1)^{n-1} \\ p_n q_{n-2} - p_{n-2} q_n &= (-1)^n a_n \end{aligned}}$$

and

$$\boxed{c_n - c_{n-1} = \frac{(-1)^{n-1}}{q_n\, q_{n-1}} \quad , \quad c_n - c_{n-2} = \frac{(-1)^n a_n}{q_n\, q_{n-2}},}$$

where $c_n - c_{n-2}$ is only valid for $n \geq 2$.

We also have the following interesting result.

Corollary 8.7. *All the p_n, q_n for a simple continued fraction are relatively prime; that is, $c_n = p_n/q_n$ is automatically in lowest terms.*

Proof. The fact that p_n, q_n are in lowest terms follows from the fact that

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1},$$

so if an integer happens to divide divide both p_n and q_n, then it divides $p_n q_{n-1} - p_{n-1} q_n$ also, so it must divide $(-1)^{n-1} = \pm 1$ which is impossible unless the number was ± 1. □

8.3.2. ★ Diophantine equations and sailors, coconuts, and monkeys.

From Section 8.1, we know that any rational number can be written as a finite simple continued fraction. Also, any finite simple continued fraction is certainly a rational number because it is made up of additions and divisions of rational numbers and the rational numbers are closed under such operations. (For proofs of these statements see Problem 2 in Exercises 8.1.) Now as we showed at the beginning of Section 8.1, we can write

$$\frac{157}{68} = 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5}}} = \langle 2; 3, 4, 5 \rangle,$$

which has an odd number of terms (three to be exact) after the integer part 2. Observe that we can write this in another way:

$$\frac{157}{68} = 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{4 + \cfrac{1}{1}}}} = \langle 2; 3, 4, 4, 1 \rangle,$$

which has an even number of terms after the integer part. This example is typical: Any finite simple continued fraction can be written with an even or odd number of terms (by modifying the last term by 1). We summarize these remarks in the following theorem, which we shall use in Theorem 8.9.

Theorem 8.8. *A real number can be expressed as a finite simple continued fraction if and only if it is rational, in which case, the rational number can be expressed as a continued fraction with either an even or an odd number of terms.*

The proof of this theorem shall be left to you. We now come to the subject of diophantine equations, which are polynomial equations we wish to find integer solutions. We shall study very elementary diophantine equations in this section, the linear ones. Before doing so, it might of interest to know that diophantine equations is named after a Greek mathematician Diophantus of Alexandrea (200 A.D. –284 A.D.). Diophantus is famous for at least two things: His book *Arithmetica*, which studies equations that we now call diophantine equations in his honor, and for the following riddle, which was supposedly written on his tombstone:

> *This tomb hold Diophantus Ah, what a marvel! And the tomb tells scientifically the measure of his life. God vouchsafed that he should be a boy for the sixth part of his life; when a twelfth was added, his cheeks acquired a beard; He kindled for him the light of marriage after a seventh, and in the fifth year after his marriage He granted him a son. Alas! late-begotten and miserable child, when he had reached the measure of half his father's life, the chill grave took him. After consoling his grief by this science of numbers for four years, he reached the end of his life.* [**160**].

Try to find how old Diophantus was when he died using elementary algebra. (Let $x =$ his age when he died; then you should end up with trying to solve the equation $x = \frac{1}{6}x + \frac{1}{12}x + \frac{1}{7}x + 5 + \frac{1}{2}x + 4$, obtaining $x = 84$.) Here is an easy way to find his age: Unravelling the above fancy language, and picking out two facts, we know that 1/12-th of his life was in youth and 1/7-th was as a bachelor. In particular, his age must divide 7 and 12. The only integer that does this, and which is within the human lifespan, is $7 \cdot 12 = 84$. In particular, he spent $84/6 = 14$ years as a child, $84/12 = 7$ as a youth, $84/7 = 12$ years as a bachelor. He married at $14 + 7 + 12 = 33$, at $33 + 5 = 38$, his son was born, who later died at the age of $84/2 = 42$ years old, when Diophantus was 80. Finally, after 4 years doing the "science of numbers", Diophantus died at the ripe old age of 84.

After taking a moment to wipe away our tears, let us consider the following.

Theorem 8.9. *If $a, b \in \mathbb{N}$ are relatively prime, then for any $c \in \mathbb{Z}$, the equation*

$$ax - by = c$$

has an infinite number of integer solutions (x, y). Moreover, if (x_0, y_0) is any one integral solution of the equation with $c = 1$, then for general $c \in \mathbb{Z}$, all solutions are of the form

$$x = cx_0 + bt \quad , \quad y = cy_0 + at \quad , \quad \textit{for all} \;\; t \in \mathbb{Z}.$$

Proof. In Problem 7 we ask you to prove this theorem using Problem 5 in Exercises 2.4; but we shall use continued fractions just for fun. We first solve the equation $ax - by = 1$. To do so, we write a/b as a finite simple continued fraction: $a/b = \langle a_0; a_1, a_2, \ldots, a_n \rangle$ and by Theorem 8.8 we can choose n to be *odd*. Then a/b is equal to the n-th convergent p_n/q_n, which implies that $p_n = a$ and $q_n = b$. Also, by our relations in Corollary 8.6, we know that

$$p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1} = 1,$$

where we used that n is odd. Since p_n and q_n are relatively prime and $a/b = p_n/q_n$, we must have $p_n = a$ and $q_n = b$. Therefore, $aq_{n-1} - bp_{n-1} = 1$, so

$$\boxed{(x_0, y_0) = (q_{n-1}, p_{n-1})} \tag{8.20}$$

solves $ax_0 - by_0 = 1$. Multiplying $ax_0 - by_0 = 1$ by c we get

$$a(cx_0) - b(cy_0) = c.$$

Then $ax - by = c$ holds if and only if

$$ax - by = a(cx_0) - b(cy_0) \quad \Longleftrightarrow \quad a(x - cx_0) = b(y - cy_0).$$

This shows that a divides $b(y - cy_0)$, which can be possible if and only if a divides $y - cy_0$ since a and b are relatively prime. Thus, $y - cy_0 = at$ for some $t \in \mathbb{Z}$. Plugging $y - cy_0 = at$ into the equation $a(x - cx_0) = b(y - cy_0)$, we get

$$a(x - cx_0) = b \cdot (at) = abt.$$

Cancelling a, we get $x - cx_0 = bt$ and our proof is now complete. □

We remark that we need a and b to be relatively prime; for example, the equation $2x - 4y = 1$ has no integer solutions (because the left hand side is always even, so can never equal 1). We also remark that the proof of Theorem 8.9, in particular, the formula (8.20), also shows us *how* to find (x_0, y_0): Just write a/b as a simple continued fraction with an *odd* number n terms after the integer part of a/b and compute the $(n-1)$-st convergent to get $(x_0, y_0) = (q_{n-1}, p_{n-1})$.

Example 8.10. Consider the diophantine equation

$$157x - 68y = 12.$$

We already know that the continued fraction expansion of $a/b = \frac{157}{68}$ with an odd $n = 3$ is $\frac{157}{68} = \langle 2; 3, 4, 5\rangle = \langle a_0; a_1, a_2, a_3\rangle$. Thus,

$$c_2 = 2 + \cfrac{1}{3 + \cfrac{1}{4}} = 2 + \frac{4}{13} = \frac{30}{13}.$$

Therefore, $(13, 30)$ is one solution of $157x - 68y = 1$, which we should check just to be sure:

$$157 \cdot 13 - 68 \cdot 30 = 2041 - 2040 = 1.$$

Since $cx_0 = 12 \cdot 13 = 156$ and $cy_0 = 12 \cdot 30 = 360$, the general solution of $157x - 68y = 12$ is

$$x = 156 + 68t \quad , \quad y = 360 + 157t, \qquad t \in \mathbb{Z}.$$

Example 8.11. We now come to a fun puzzle that involves diophantine equations; for more cool coconut puzzles, see [**80, 81**], [**228**], [**212**], and Problem 5. See also [**214**] for the long history of such problems. Five sailors get shipwrecked on an island where there is only a coconut tree and a very slim monkey. The sailors gathered all the coconuts into a gigantic pile and went to sleep. At midnight, one sailor woke up, and because he didn't trust his mates, he divided the coconuts into five equal piles, but with one coconut left over. He throws the extra one to the monkey, hides his pile, puts the remaining coconuts back into a pile, and goes to sleep. At one o'clock, the second sailor woke up, and because he was untrusting of his mates, he divided the coconuts into five equal piles, but again with one coconut left over. He throws the extra one to the monkey, hides his pile, puts the remaining coconuts back into a pile, and goes to sleep. This exact same scenario continues throughout the night with the other three sailors. In the morning, all the sailors woke up, pretending as if nothing happened and divided the now minuscule pile of coconuts into five equal piles, and they find yet again one coconut left over, which they throw to the now very overweight monkey. Question: What is the smallest possible number of coconuts in the original pile?

Let $x =$ the original number of coconuts. Remember that sailor #1 divided x into five parts, but with one left over. This means that if y_1 is the number that he took, then $x = 5y_1 + 1$ and he left $4 \cdot y_1$ coconuts. In other words, he took

$$\frac{1}{5}(x-1) \text{ coconuts, leaving } 4 \cdot \frac{1}{5}(x-1) = \frac{4}{5}(x-1) \text{ coconuts.}$$

Similarly, if y_2 is the number of coconuts that sailor #2 took, then $\frac{4}{5}(x-1) = 5y_2+1$ and he left $4 \cdot y_2$ coconuts. That is, the second sailor took

$$\frac{1}{5} \cdot \left(\frac{4}{5}(x-1) - 1\right) = \frac{4x-9}{25} \text{ coconuts, leaving } 4 \cdot \frac{4x-9}{25} = \frac{16x-36}{25} \text{ coconuts.}$$

Repeating this argument, we find that sailors #3, #4, and #5 left

$$\frac{64x-244}{125}\ ,\ \frac{256x-1476}{625}\ ,\ \frac{1024x-8404}{3125}$$

coconuts, respectively. At the end, the sailors divided this last amount of coconuts into five piles, with one coconut left over. Thus, if y is the number of coconuts in each pile, then we must have

$$\frac{1024x-8404}{3125} = 5y+1 \quad \Longrightarrow \quad 1024x - 15625y = 11529.$$

The equation $1024x - 15625y = 11529$ is just a diophantine equation since we want *integers* x, y solving this equation. Moreover, $1024 = 2^{10}$ and $15625 = 5^6$ are relatively prime, so we can solve this equation by Theorem 8.9. First, we solve $1024x - 15625y = 1$ by writing $1024/15625$ as a continued fraction (this takes some algebra) and forcing n to be odd (in this case $n = 9$):[2]

$$\frac{1024}{15625} = \langle 0; 15, 3, 1, 6, 2, 1, 3, 2, 1\rangle.$$

Second, we take the $(n-1)$-st convergent:

$$c_8 = \langle 0; 15, 3, 1, 6, 2, 1, 3, 2\rangle = \frac{711}{10849}.$$

Thus, $(x_0, y_0) = (10849, 711)$. Since $cx_0 = 11529 \cdot 10849 = 125078121$ and $cy_0 = 11529 \cdot 711 = 8197119$, the integer solutions to $1024x - 15625y = 11529$ are

$$x = 125078121 + 15625t \quad , \quad y = 8197119 + 1024t \quad , \quad t \in \mathbb{Z}. \tag{8.21}$$

This of course gives us infinitely many solutions. However, we want the smallest *nonnegative* solutions since x and y represent numbers of coconuts; thus, we need

$$x = 125078121 + 15625t \geq 0 \quad \Longrightarrow \quad t \geq -\frac{125078121}{15625} = -8004.999744\ldots,$$

and

$$y = 8197119 + 1024t \geq 0 \quad \Longrightarrow \quad t \geq -\frac{8197119}{1024} = -8004.9990234\ldots.$$

Thus, taking $t = -8004$ in (8.21), we finally arrive at $x = 15621$ and $y = 1023$. In conclusion, the smallest number of coconuts in the original piles is 15621 coconuts. By the way, you can solve this coconut problem *without* continued fractions using nothing more than basic high school algebra; try it!

Exercises 8.3.

1. Find the general integral solutions of

$$(a)\ 7x - 11y = 1 \quad , \quad (b)\ 13x - 3y = 5 \quad , \quad (c)\ 13x - 15y = 100.$$

[2] See http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/cfCALC.html for a continued fraction calculator.

2. If all the $a_0, a_1, a_2, \ldots, a_n > 0$ (which guarantees that $p_0 = a_0 > 0$), prove that
$$\frac{p_n}{p_{n-1}} = \langle a_n; a_{n-1}, a_{n-2}, \ldots, a_2, a_1, a_0\rangle \quad \text{and} \quad \frac{q_n}{q_{n-1}} = \langle a_n; a_{n-1}, a_{n-2}, \ldots, a_2, a_1\rangle$$
for $n = 1, 2, \ldots$. Suggestion: Observe that $\frac{p_k}{p_{k-1}} = \frac{a_k p_{k-1} + p_{k-2}}{p_{k-1}} = a_k + \frac{1}{\frac{p_{k-1}}{p_{k-2}}}$.
3. In this problem, we relate the Fibonacci numbers to continued fractions. Recall that the Fibonacci sequence $\{F_n\}$ is defined as $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$. Let $p_n/q_n = \langle a_0; a_1, \ldots, a_n\rangle$ where all the a_k's are equal to 1.
   (a) Prove that $p_n = F_{n+2}$ and $q_n = F_{n+1}$ for all $n = -1, 0, 1, 2, \ldots$. Suggestion: Use the Wallis-Euler recurrence relations.
   (b) From facts known about convergents, prove that F_n and F_{n+1} are relatively prime and derive the following famous identity, named after Giovanni Domenico Cassini (1625–1712) (also called Jean-Dominique Cassini)
$$F_{n-1}F_{n+1} - F_n^2 = (-1)^n \qquad (\textbf{Cassini's identity}).$$
4. Imitating the proof of Theorem 8.9, show that a solution of $ax - by = -1$ can be found by writing a/b as a simple continued fraction with an *even* number n terms after the integer part of a/b and finding the $(n-1)$-th convergent. Apply this method to find a solution of $157x - 68y = -1$ and $7x - 12y = -1$.
5. (**Coconut problems**) Here are some more coconut problems:
   (a) Solve the coconut problem assuming the same antics as in the text, except for one thing: there are no coconuts left over for the monkey at the end. That is, what is the smallest possible number of coconuts in the original pile given that after the sailors divided the coconuts in the morning, there are no coconuts left over?
   (b) Solve the coconut problem assuming the same antics as in the text except that during the night each sailor divided the pile into five equal piles with *none* left over; however, after he puts the remaining coconuts back into a pile, the monkey (being a thief himself) steals one coconut from the pile (before the next sailor wakes up). In the morning, there is still one coconut left over for the monkey.
   (c) Solve the coconut problem when there are only three sailors to begin with, otherwise everything is the same as in the text (e.g. one coconut left over at the end). Solve this same coconut problem when there are no coconuts left over at the end.
   (d) Solve the coconut problem when there are seven sailors, otherwise everything is the same as in the text. (Warning: Set aside an evening for long computations!)
6. Let $\alpha = \langle a_0; a_1, a_2, \ldots, a_m\rangle$, $\beta = \langle b_0; b_1, \ldots, b_n\rangle$ with $m, n \geq 0$ and the a_k, b_k's integers with $a_m, b_n > 1$ (such finite continued fractions are called **regular**). Prove that if $\alpha = \beta$, then $a_k = b_k$ for all $k = 0, 1, 2, \ldots$. In other words, distinct regular finite simple continued fractions define different rational numbers.
7. Prove Theorem 8.9 using Problem 5 in Exercises 2.4.

8.4. Convergence theorems for infinite continued fractions

Certainly the continued fraction $\langle 1; 1, 1, 1, 1, \ldots\rangle$ (if it converges), should be a very special number — it is, it turns out to be the golden ratio! In this section we shall investigate the delicate issues surrounding convergence of infinite continued fractions (see Theorem 8.14, the continued fraction convergence theorem); in particular, we prove that *any* simple continued fraction converges. We also show how to write *any* real number as a simple continued fraction via the **canonical continued fraction algorithm**. Finally, we prove that a real number is irrational if and only if its simple continued fraction expansion is infinite.

8.4.1. Monotonicity properties of convergents. Let $\{c_n\}$ denote the convergents of a nonnegative infinite continued fraction

$$a_0 + \frac{b_1}{a_1 +} \frac{b_2}{a_2 +} \frac{b_3}{a_3 +} \cdots \frac{b_n}{+\, a_n +} \cdots,$$

where recall that nonnegative means the a_n, b_n's are real numbers with $a_n > 0, b_n \geq 0$ for all $n \geq 1$, and there is no restriction on a_0. The Wallis-Euler recurrence relations (8.16) are

$$\begin{aligned} p_n &= a_n p_{n-1} + b_n p_{n-2} \quad , \quad q_n = a_n q_{n-1} + b_n q_{n-2} \\ p_{-1} &= 1 \ , \ p_0 = a_0 \quad , \quad q_{-1} = 0 \ , \ q_0 = 1. \end{aligned}$$

Then (cf. (8.17))

$$p_1 = a_1 p_0 + b_1 p_{-1} = a_1 a_0 + b_1 \ \ , \ \ q_1 = a_1 q_0 + b_1 q_{-1} = a_1,$$

and all the q_n's are positive (see discussion below (8.17)). By Theorem 8.4 we have $c_n = p_n/q_n$ for all n and by Theorem 8.5, for all $n \geq 1$ the fundamental recurrence relations are

$$\begin{aligned} p_n q_{n-1} - p_{n-1} q_n &= (-1)^{n-1} b_1 b_2 \cdots b_n \\ p_n q_{n-2} - p_{n-2} q_n &= (-1)^n a_n b_1 b_2 \cdots b_{n-1} \end{aligned}$$

and

$$c_n - c_{n-1} = \frac{(-1)^{n-1} b_1 b_2 \cdots b_n}{q_n \, q_{n-1}} \quad , \quad c_n - c_{n-2} = \frac{(-1)^n a_n b_1 b_2 \cdots b_{n-1}}{q_n \, q_{n-2}},$$

where $c_n - c_{n-2}$ is only valid for $n \geq 2$. Using these fundamental recurrence relations, we shall prove the following monotonicity properties of the c_n's, which is important in the study of the convergence properties of the c_n's.

Theorem 8.10. *Assume that $b_n > 0$ for each n. Then the convergents $\{c_n\}$ satisfy the inequalities: For all $n \in \mathbb{N}$,*

$$c_0 < c_2 < c_4 < \cdots < c_{2n} < c_{2n-1} < \cdots < c_5 < c_3 < c_1.$$

That is, the even convergents form a strictly increasing sequence while the odd convergent form a strictly decreasing sequence.

Proof. Replacing n with $2n$ in the fundamental recurrence relation for $c_n - c_{n-2}$, we see that

$$c_{2n} - c_{2n-2} = \frac{(-1)^{2n-2} a_{2n} b_1 b_2 \cdots b_{2n-1}}{q_{2n} \, q_{2n-1}} = \frac{a_{2n} b_1 b_2 \cdots b_{2n-1}}{q_{2n} \, q_{2n-1}} > 0.$$

This shows that $c_{2n-2} < c_{2n}$ for all $n \geq 1$ and hence, $c_0 < c_2 < c_4 < \cdots$. Replacing n with $2n - 1$ in the fundamental relation for $c_n - c_{n-2}$, one can prove that the odd convergents form a strictly decreasing sequence. Replacing n with $2n$ in the fundamental recurrence relation for $c_n - c_{n-1}$, we see that

$$c_{2n} - c_{2n-1} = \frac{(-1)^{2n-1} b_1 b_2 \cdots b_{2n}}{q_{2n} \, q_{2n-1}} = -\frac{b_1 b_2 \cdots b_{2n}}{q_{2n} \, q_{2n-1}} < 0 \implies c_{2n} < c_{2n-1}. \tag{8.22}$$

□

If the continued fraction is actually finite; that is, if $b_{\ell+1} = 0$ for some ℓ, then this theorem still holds, but we need to make sure that $2n \leq \ell$. By the monotone criterion for sequences, we have

COROLLARY 8.11. *The limits of the even and odd convergents exist, and*

$$c_0 < c_2 < c_4 < \cdots < \lim c_{2n} \le \lim c_{2n-1} < \cdots < c_5 < c_3 < c_1.$$

8.4.2. Convergence results for continued fractions. As a consequence of the previous corollary, it follows that $\lim c_n$ exists if and only if $\lim c_{2n} = \lim c_{2n-1}$, which holds if and only if

$$c_{2n} - c_{2n-1} = \frac{-b_1 b_2 \cdots b_{2n}}{q_{2n}\, q_{2n-1}} \to 0 \quad \text{as } n \to \infty. \tag{8.23}$$

In the following theorem, we give one condition under which this is satisfied.

THEOREM 8.12. *Let $\{a_n\}_{n=0}^\infty, \{b_n\}_{n=1}^\infty$ be sequences such that $a_n, b_n > 0$ for $n \ge 1$ and*

$$\sum_{n=1}^\infty \frac{a_n a_{n+1}}{b_{n+1}} = \infty.$$

Then (8.23) holds, so the continued fraction $\xi := a_0 + \frac{b_1}{a_1 +} \frac{b_2}{a_2 +} \frac{b_3}{a_3 +} \frac{b_4}{a_4 +} \cdots$ converges. Moreover, for any even j and odd k, we have

$$c_0 < c_2 < c_4 < \cdots < c_j < \cdots < \xi < \cdots < c_k < \cdots < c_5 < c_3 < c_1.$$

PROOF. Observe that for any $n \ge 2$, we have $q_{n-1} = a_{n-1} q_{n-2} + b_{n-1} q_{n-3} \ge a_{n-1} q_{n-2}$ since $b_{n-1}, q_{n-3} \ge 0$. Thus, for $n \ge 2$ we have

$$q_n = a_n q_{n-1} + b_n q_{n-2} \ge a_n \cdot (a_{n-1} q_{n-2}) + b_n q_{n-2} = q_{n-2}(a_n a_{n-1} + b_n),$$

so

$$q_n \ge q_{n-2}(a_n a_{n-1} + b_n).$$

Applying this formula over and over again, we find that for any $n \ge 1$,

$$\begin{aligned} q_{2n} &\ge q_{2n-2}(a_{2n} a_{2n-1} + b_{2n}) \\ &\ge q_{2n-4}(a_{2n-2} a_{2n-3} + b_{2n-2}) \cdot (a_{2n} a_{2n-1} + b_{2n}) \\ &\ge \ \vdots \\ &\ge q_0 (a_2 a_1 + b_2)(a_4 a_3 + b_4) \cdots (a_{2n} a_{2n-1} + b_{2n}). \end{aligned}$$

A similar argument shows that for any $n \ge 2$,

$$q_{2n-1} \ge q_1 (a_3 a_2 + b_3)(a_5 a_4 + b_5) \cdots (a_{2n-1} a_{2n-2} + b_{2n-1}).$$

Thus, for any $n \ge 2$, we have

$$q_{2n} q_{2n-1} \ge q_0 q_1 (a_2 a_1 + b_2)(a_3 a_2 + b_3) \cdots (a_{2n-1} a_{2n-2} + b_{2n-1})(a_{2n} a_{2n-1} + b_{2n}).$$

Factoring out all the b_k's we conclude that

$$q_{2n} q_{2n-1} \ge q_0 q_1 b_2 \cdots b_{2n} \cdots \left(1 + \frac{a_2 a_1}{b_2}\right)\left(1 + \frac{a_3 a_2}{b_3}\right) \cdots \left(1 + \frac{a_{2n} a_{2n-1}}{b_{2n}}\right),$$

which shows that

$$\frac{b_1 b_2 \cdots b_{2n}}{q_{2n}\, q_{2n-1}} \le \frac{b_1}{q_0 q_1} \cdot \frac{1}{\prod_{k=1}^{2n-1}\left(1 + \frac{a_k a_{k+1}}{b_{k+1}}\right)}. \tag{8.24}$$

Now recall that (see Theorem 7.2) a series $\sum_{k=1}^\infty \alpha_k$ of positive real numbers converges if and only if the infinite product $\prod_{k=1}^\infty (1 + \alpha_k)$ converges. Thus, since we are given that $\sum_{k=1}^\infty \frac{a_k a_{k+1}}{b_{k+1}} = \infty$, we have $\prod_{k=1}^\infty \left(1 + \frac{a_k a_{k+1}}{b_{k+1}}\right) = \infty$ as well, so the right-hand side of (8.24) vanishes as $n \to \infty$. The fact that for even j and odd k, we have $c_0 < c_2 < c_4 < \cdots < c_j < \cdots < \xi < \cdots < c_k < \cdots < c_5 < c_3 < c_1$ follows from Corollary 8.11. This completes our proof. □

For another convergence theorem, see Problems 6 and 9. An important example for which this theorem applies is to simple continued fractions: For a simple continued fraction $\langle a_0; a_1, a_2, a_3, \ldots \rangle$, all the b_n's equal 1, so

$$\sum_{n=1}^{\infty} \frac{a_n a_{n+1}}{b_{n+1}} = \sum_{n=1}^{\infty} a_n a_{n+1} = \infty,$$

since all the a_n's are positive integers. Thus,

Corollary 8.13. *Infinite simple continued fractions always converge and if ξ is the limit of such a fraction, then for any even j and odd k, the convergents satisfy*

$$c_0 < c_2 < c_4 < \cdots < c_j < \cdots < \xi < \cdots < c_k < \cdots < c_5 < c_3 < c_1.$$

Example 8.12. In particular, the very special fraction $\Phi := \langle 1; 1, 1, 1, \ldots \rangle$ converges. To what you ask? Observe that

$$\Phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}} = 1 + \frac{1}{\Phi} \quad \Longrightarrow \quad \Phi = 1 + \frac{1}{\Phi}.$$

We can also get this formula from convergents: The n-th convergent of Φ is

$$c_n = 1 + \cfrac{1}{1 + \cfrac{1}{\ddots \, 1 + \cfrac{1}{1 + \cfrac{1}{1}}}} = 1 + \frac{1}{c_{n-1}}.$$

Thus, if we set $\Phi = \lim c_n$, which we know exists, then taking $n \to \infty$ on both sides of $c_n = 1 + \frac{1}{c_{n-1}}$, we get $\Phi = 1 + 1/\Phi$ just as before. Thus, $\Phi^2 - \Phi - 1 = 0$, which after solving for Φ we get

$$\Phi = \frac{1 + \sqrt{5}}{2},$$

the golden ratio.

As a unrelated side note, we remark that Φ can be used to get a fairly accurate (and well-known) approximation to π:

$$\boxed{\pi \approx \frac{6}{5}\Phi^2 = 3.1416\ldots.}$$

Example 8.13. The continued fraction $\xi := 3 + \frac{4}{6} {}_{+} \frac{4}{6} {}_{+} \frac{4}{6} {}_{+} \frac{4}{6} \cdots$ was studied by Rafael Bombelli (1526–1572) and was one of the first continued fractions ever to be studied. Since $\sum_{n=1}^{\infty} \frac{a_n a_{n+1}}{b_{n+1}} = \sum_{n=1}^{\infty} \frac{6^2}{4} = \infty$, this continued fraction converges.

By the same reasoning, the continued fraction $\eta := 6 + \frac{4}{6} {}_{+} \frac{4}{6} {}_{+} \frac{4}{6} \ldots$ also converges. Moreover, $\xi = \eta - 3$ and

$$\eta = 6 + \cfrac{4}{6 + \cfrac{4}{6 + \cfrac{4}{\ddots}}} = 6 + \frac{1}{\eta} \quad \Longrightarrow \quad \eta = 6 + \frac{1}{\eta} \quad \Longrightarrow \quad \eta^2 - 6\eta - 1 = 0.$$

Solving this quadratic equation for η, we find that $\eta = 3 + \sqrt{13}$. Hence, $\xi = \eta - 3 = \sqrt{13}$. Isn't this fun!

8.4.3. The canonical continued fraction algorithm and the continued fraction convergence theorem. What if we want to *construct* the continued fraction expansion of a real number? We know how to construct such an expansion for rational numbers, so let us review this; the same method will work for irrational numbers. Consider again our friend $\frac{157}{68} = \langle 2; 3, 4, 5 \rangle = \langle a_0; a_1, a_2, a_3 \rangle$, and let us recall how we found its continued fraction expansion. First, we wrote $\xi_0 := \frac{157}{68}$ as

$$\xi_0 = 2 + \frac{1}{\xi_1} \quad , \quad \text{where } \xi_1 = \frac{68}{21} > 1.$$

In particular, notice that

$$a_0 = 2 = \lfloor \xi_0 \rfloor,$$

where recall that$\lfloor x \rfloor$, where x is a real number, denotes the largest integer $\leq x$. Second, we wrote $\xi_1 = \frac{68}{21}$ as

$$\xi_1 = 3 + \frac{1}{\xi_2} \quad , \quad \text{where } \xi_2 = \frac{21}{5} > 1.$$

In particular, notice that

$$a_1 = 3 = \lfloor \xi_1 \rfloor.$$

Third, we wrote

$$\xi_2 = \frac{21}{5} = 4 + \frac{1}{\xi_3} \quad , \quad \text{where } \xi_3 = 5 > 1.$$

In particular, notice that

$$a_2 = 4 = \lfloor \xi_2 \rfloor.$$

Finally, $a_3 = \lfloor \xi_3 \rfloor = \xi_3$ cannot be broken up any further so we *stop here*. Hence,

$$\frac{157}{68} = \xi_0 = 2 + \frac{1}{\xi_1} = 2 + \cfrac{1}{3 + \cfrac{1}{\xi_2}} = 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{\xi_3}}} = 2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{5}}}.$$

We've just found the **canonical (simple) continued fraction** of 157/68. Notice that we end with the number 5, which is greater than 1; this will always happen whenever we do the above procedure for a noninteger rational number (such continued fractions were called **regular** in Problem 6 of Exercises 8.3). We can do the same exact procedure for irrational numbers! Let ξ be an irrational number. First, we set $\xi_0 = \xi$ and define $a_0 := \lfloor \xi_0 \rfloor \in \mathbb{Z}$. Then, $0 < \xi_0 - a_0 < 1$ (note that $\xi_0 \neq a_0$ since ξ_0 is irrational), so we can write

$$\xi_0 = a_0 + \frac{1}{\xi_1} \quad , \quad \text{where } \xi_1 := \frac{1}{\xi_0 - a_0} > 1,$$

where we used that $0 < \xi_0 - a_0$. Note that ξ_1 is irrational because if not, then ξ_0 would be rational contrary to assumption. Second, we define $a_1 := \lfloor \xi_1 \rfloor \in \mathbb{N}$. Then, $0 < \xi_1 - a_1 < 1$, so we can write

$$\xi_1 = a_1 + \frac{1}{\xi_2} \quad , \quad \text{where } \xi_2 := \frac{1}{\xi_1 - a_1} > 1.$$

Note that ξ_2 is irrational. Third, we define $a_2 := \lfloor \xi_2 \rfloor \in \mathbb{N}$. Then, $0 < \xi_2 - a_2 < 1$, so we can write

$$\xi_2 = a_2 + \frac{1}{\xi_3} \quad , \quad \text{where } \xi_3 := \frac{1}{\xi_2 - a_2} > 1.$$

Note that ξ_3 is irrational. We can continue this procedure to "infinity" creating a sequence $\{\xi_n\}_{n=0}^\infty$ of real numbers with $\xi_n > 0$ for $n \geq 1$ called the **complete quotients** of ξ, and a sequence $\{a_n\}_{n=0}^\infty$ of integers with $a_n > 0$ for $n \geq 1$ called the **partial quotients** of ξ, such that

$$\xi_n = a_n + \frac{1}{\xi_{n+1}}, \quad n = 0, 1, 2, 3, \ldots.$$

Thus,

$$\xi = \xi_0 = a_0 + \frac{1}{\xi_1} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\xi_2}} = \cdots \text{``}=\text{''}\, a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \ddots}}}}. \tag{8.25}$$

We emphasize that we have actually not proved that ξ is *equal* to the infinite continued fraction on the far right (hence, the quotation marks)! But, as a consequence of the following theorem, this equality follows; then the continued fraction in (8.25) is called the **canonical (simple) continued fraction expansion** of ξ.

THEOREM 8.14 (**Continued fraction convergence theorem**). *Let ξ_0, ξ_1, ξ_2, ... be any sequence of real numbers with $\xi_n > 0$ for $n \geq 1$ and suppose that these numbers are related by*

$$\xi_n = a_n + \frac{b_{n+1}}{\xi_{n+1}} \quad , \quad n = 0, 1, 2, \ldots,$$

for sequences of real numbers $\{a_n\}_{n=0}^\infty, \{b_n\}_{n=1}^\infty$ with $a_n, b_n > 0$ for $n \geq 1$ and which satisfy $\sum_{n=1}^\infty \frac{a_n a_{n+1}}{b_{n+1}} = \infty$. Then ξ_0 is equal to the continued fraction

$$\xi_0 = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\frac{b_4}{a_4 +}\,\frac{b_5}{a_5 +}\,\cdots.$$

In particular, for any real number ξ, the canonical continued fraction expansion (8.25) converges to ξ.

PROOF. By Theorem 8.12, the continued fraction $a_0 + \frac{b_1}{a_1 +}\frac{b_2}{a_2 +}\frac{b_3}{a_3 +}\cdots$ converges. Let $\{c_k = p_k/q_k\}$ denote the convergents of this infinite continued fraction and let $\varepsilon > 0$. Then by Theorem 8.12, there is an N such that

$$n > N \implies |c_n - c_{n-1}| = \frac{b_1 b_2 \cdots b_n}{q_n\, q_{n-1}} < \varepsilon.$$

Fix $n > N$ and consider the *finite* continued fraction obtained as in (8.25) by writing out ξ_0 to the n-th term:

$$\xi_0 = a_0 + \frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\,\frac{b_3}{a_3+}\,\cdots\,\frac{b_{n-1}}{+\,a_{n-1}+}\,\frac{b_n}{\xi_n}.$$

Let $\{c'_k = p'_k/q'_k\}$ denote the convergents of this finite continued fraction. Then observe that $p_k = p'_k$ and $q_k = q'_k$ for $k \le n-1$ and $c'_n = \xi_0$. Therefore, by our fundamental recurrence relations, we have

$$|\xi_0 - c_{n-1}| = |c'_n - c'_{n-1}| \le \frac{b_1 b_2 \cdots b_n}{q'_n\, q'_{n-1}} = \frac{b_1 b_2 \cdots b_n}{q'_n\, q_{n-1}}.$$

By the Wallis-Euler relations, we have

$$q'_n = \xi_n q'_{n-1} + b_n q'_{n-2} = \left(a_n + \frac{b_{n+1}}{\xi_{n+1}}\right) q_{n-1} + b_n q_{n-2} > a_n q_{n-1} + b_n q_{n-2} = q_n.$$

Hence,

$$|\xi_0 - c_{n-1}| \le \frac{b_1 b_2 \cdots b_n}{q'_n\, q_{n-1}} < \frac{b_1 b_2 \cdots b_n}{q_n\, q_{n-1}} < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\xi_0 = \lim c_{n-1} = \xi$. □

Example 8.14. Consider $\xi_0 = \sqrt{3} = 1.73205\ldots$. In this case, $a_0 := \lfloor \xi_0 \rfloor = 1$. Thus,

$$\xi_1 := \frac{1}{\xi_0 - a_0} = \frac{1}{\sqrt{3}-1} = \frac{1+\sqrt{3}}{2} = 1.36602\ldots \quad \Longrightarrow \quad a_1 := \lfloor \xi_1 \rfloor = 1.$$

Therefore,

$$\xi_2 := \frac{1}{\xi_1 - a_1} = \frac{1}{\dfrac{1+\sqrt{3}}{2} - 1} = 1 + \sqrt{3} = 2.73205\ldots \quad \Longrightarrow \quad a_2 := \lfloor \xi_2 \rfloor = 2.$$

Hence,

$$\xi_3 := \frac{1}{\xi_2 - a_2} = \frac{1}{\sqrt{3}-1} = \frac{1+\sqrt{3}}{2} = 1.36602\ldots \quad \Longrightarrow \quad a_3 := \lfloor \xi_3 \rfloor = 1.$$

Here we notice that $\xi_3 = \xi_1$ and $a_3 = a_1$. Therefore,

$$\xi_4 := \frac{1}{\xi_3 - a_3} = \frac{1}{\xi_1 - a_1} = \xi_2 = 1 + \sqrt{3} \quad \Longrightarrow \quad a_4 := \lfloor \xi_4 \rfloor = \lfloor \xi_2 \rfloor = 2.$$

At this point, we see that we will get the repeating pattern $1, 2, 1, 2, \ldots$, so we conclude that

$$\sqrt{3} = \langle 1; 1, 2, 1, 2, 1, 2, \ldots \rangle = \langle 1; \overline{1, 2} \rangle,$$

where we indicate that the $1, 2$ pattern repeats by putting a bar over them.

Example 8.15. Here is a neat example concerning the Fibonacci and Lucas numbers; for other fascinating topics on these numbers, see Knott's fun website [**121**]. Let us find the continued fraction expansion of the irrational number $\xi_0 = \Phi/\sqrt{5}$ where Φ is the golden ratio $\Phi = \frac{1+\sqrt{5}}{2}$:

$$\xi_0 = \frac{\Phi}{\sqrt{5}} = \frac{1+\sqrt{5}}{2\sqrt{5}} = 0.72360679\ldots \quad \Longrightarrow \quad a_0 := \lfloor \xi_0 \rfloor = 0.$$

Thus,

$$\xi_1 := \frac{1}{\xi_0 - a_0} = \frac{1}{\xi_0} = \frac{2\sqrt{5}}{1+\sqrt{5}} = 1.3819660\ldots \quad \Longrightarrow \quad a_1 := \lfloor \xi_1 \rfloor = 1.$$

Therefore,

$$\xi_2 := \frac{1}{\xi_1 - a_1} = \frac{1}{\dfrac{2\sqrt{5}}{1+\sqrt{5}} - 1} = \frac{1+\sqrt{5}}{\sqrt{5}-1} = 2.6180339\ldots \quad \Longrightarrow \quad a_2 := \lfloor \xi_2 \rfloor = 2.$$

Hence,

$$\xi_3 := \frac{1}{\xi_2 - a_2} = \frac{1}{\dfrac{1+\sqrt{5}}{\sqrt{5}-1} - 2} = \frac{\sqrt{5}-1}{3-\sqrt{5}} = 1.2360679\ldots \quad \Longrightarrow \quad a_3 := \lfloor \xi_3 \rfloor = 1.$$

Thus,

$$\xi_4 := \frac{1}{\xi_3 - a_3} = \frac{1}{\dfrac{\sqrt{5}-1}{3-\sqrt{5}} - 1} = \frac{3-\sqrt{5}}{2\sqrt{5}-4} = \frac{1+\sqrt{5}}{2} = 1.6180339\ldots;$$

that is, $\xi_4 = \Phi$, and so, $a_4 := \lfloor \xi_4 \rfloor = 1$. Let us do this one more time:

$$\xi_5 := \frac{1}{\xi_4 - a_4} = \frac{1}{\dfrac{1+\sqrt{5}}{2} - 1} = \frac{2}{\sqrt{5}-1} = \frac{1+\sqrt{5}}{2} = \Phi,$$

and so, $a_5 = a_4 = 1$. Continuing on this process, we will get $\xi_n = \Phi$ and $a_n = 1$ for the rest of the n's. In conclusion, we have

$$\frac{\Phi}{\sqrt{5}} = \langle 0; 1, 2, 1, 1, 1, 1, \ldots \rangle = \langle 0; 1, 2, \overline{1} \rangle.$$

The convergents of this continued fraction are fascinating. Recall that the Fibonacci sequence $\{F_n\}$, named after Leonardo Pisano Fibonacci (1170–1250), is defined as $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 2$, which gives the sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots.$$

The **Lucas numbers** $\{L_n\}$, named after François Lucas (1842–1891), are defined by

$$L_0 := 2\ ,\ L_1 = 1\ ,\quad L_n = L_{n-1} + L_{n-2}\ \ ,\ n = 2, 3, 4, \ldots,$$

and which give the sequence

$$2, 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, \ldots$$

If you work out the convergents of $\frac{\Phi}{\sqrt{5}} = \langle 0; 1, 2, 1, 1, 1, 1, \ldots \rangle$ what you get is the fascinating result:

$$\boxed{\begin{array}{c} \dfrac{\Phi}{\sqrt{5}} = \langle 0; 1, 2, \overline{1} \rangle \ \text{ has convergents} \\ \dfrac{0}{2}, \dfrac{1}{1}, \dfrac{2}{3}, \dfrac{3}{4}, \dfrac{5}{7}, \dfrac{8}{11}, \dfrac{13}{18}, \dfrac{21}{29}, \dfrac{34}{47}, \dfrac{55}{76}, \dfrac{89}{123}, \ldots = \dfrac{\text{Fibonacci numbers}}{\text{Lucas numbers}}; \end{array}} \tag{8.26}$$

of course, we do miss the other 1 in the Fibonacci sequence. For more fascinating facts on Fibonacci numbers see Problem 7. Finally, we remark that the canonical simple fraction expansion of a real number is unique, see Problem 8.

8.4.4. The numbers π and e. We now discuss the continued fraction expansions for the famous numbers π and e. Consider π first:

$$\xi_0 = \pi = 3.141592653\ldots \quad \Longrightarrow \quad a_0 := \lfloor \xi_0 \rfloor = 3.$$

Thus,

$$\xi_1 := \frac{1}{\pi - 3} = \frac{1}{0.141592653\ldots} = 7.062513305\ldots \quad \Longrightarrow \quad a_1 := \lfloor \xi_1 \rfloor = 7.$$

Therefore,

$$\xi_2 := \frac{1}{\xi_1 - a_1} = \frac{1}{0.062513305\ldots} = 15.99659440\ldots \quad \Longrightarrow \quad a_2 := \lfloor \xi_2 \rfloor = 15.$$

Hence,

$$\xi_3 := \frac{1}{\xi_2 - a_2} = \frac{1}{0.996594407\ldots} = 1.00341723\ldots \quad \Longrightarrow \quad a_3 := \lfloor \xi_3 \rfloor = 1.$$

Let us do this one more time:

$$\xi_4 := \frac{1}{\xi_3 - a_3} = \frac{1}{0.003417231\ldots} = 292.6345908\ldots \quad \Longrightarrow \quad a_4 := \lfloor \xi_4 \rfloor = 292.$$

Continuing this process (at Davis' Broadway cafe and after 314 free refills), we get

$$\pi = \langle 3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, 2, 2, 1, 84, 2, 1, \ldots \rangle. \tag{8.27}$$

Unfortunately (or perhaps fortunately) there is no known pattern that the partial quotients follow! The first few convergents for $\pi = 3.141592653\ldots$ are

$$c_0 = 3\ ,\ c_1 = \frac{22}{7} = 3.142857142\ldots\ ,\ c_2 = \frac{333}{106} = 3.141509433\ldots$$
$$c_4 = \frac{355}{113} = 3.141592920\ldots\ ,\ c_5 = \frac{103993}{33102} = 3.141592653\ldots.$$

In stark contrast to π, Euler's number e has a shockingly simple pattern, which we ask you to work out in Problem 2:

$$e = \langle 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots \rangle$$

We will prove that this pattern continues in Section 8.7!

8.4.5. Irrationality. We now discuss when continued fractions represent irrational numbers (cf. [**154**]).

THEOREM 8.15. *Let $\{a_n\}_{n=0}^\infty$, $\{b_n\}_{n=1}^\infty$ be sequences rational numbers such that $a_n, b_n > 0$ for $n \geq 1$, $0 < b_n \leq a_n$ for all n sufficiently large, and $\sum_{n=1}^\infty \frac{a_n a_{n+1}}{b_{n+1}} = \infty$. Then the real number*

$$\xi = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\frac{b_4}{a_4 +}\,\frac{b_5}{a_5 +}\ \cdots \quad \textit{is irrational.}$$

PROOF. First of all, the continued fraction defining ξ converges by Theorem 8.12. Suppose that $0 < b_n \leq a_n$ for all $n \geq m + 1$ with $m > 0$. Observe that if we define

$$\eta = a_m + \frac{b_{m+1}}{a_{m+1} +}\,\frac{b_{m+2}}{a_{m+2} +}\,\frac{b_{m+3}}{a_{m+3} +}\ \cdots,$$

which also converges by Theorem 8.12, then $\eta > a_m > 0$ and we can write

$$\xi = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\ \cdots\ \frac{b_m}{+\ \eta}.$$

By Theorem 8.4, we know that

$$\xi = a_0 + \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\cdots\,\frac{b_m}{+\;\eta} = \frac{\eta p_m + b_m p_{m-1}}{\eta q_m + b_m q_{m-1}}.$$

Solving the last equation for η, we get

$$\xi = \frac{\eta p_m + b_m p_{m-1}}{\eta q_m + b_m q_{m-1}} \quad\Longleftrightarrow\quad \eta = \frac{\xi b_m q_{m-1} - b_m p_{m-1}}{p_m - \xi q_m}.$$

Note that since $\eta > a_m$, we have $\xi \neq p_m/q_m$. Since all the a_n, b_n's are rational, it follows that ξ is irrational if and only if η is irrational. Thus, all we have to do is prove that η is irrational. Since a_m is rational, all we have to do is prove that $\frac{b_{m+1}}{a_{m+1} +}\,\frac{b_{m+2}}{a_{m+2} +}\,\frac{b_{m+3}}{a_{m+3} +}\,\cdots$ is irrational, where $0 < b_n \le a_n$ for all $n \ge m+1$. In conclusion, we might as well assume from the start that

$$\xi = \frac{b_1}{a_1 +}\,\frac{b_2}{a_2 +}\,\frac{b_3}{a_3 +}\,\frac{b_4}{a_4 +}\,\frac{b_5}{a_5 +}\,\cdots$$

where $0 < b_n \le a_n$ for all n. We shall do this for the rest of the proof. Assume, by way of contradiction, that ξ *is* rational. Define $\xi_n := \frac{b_n}{a_n +}\,\frac{b_{n+1}}{a_{n+1} +}\,\frac{b_{n+2}}{a_{n+2} +}\,\cdots$. Then for each $n = 1, 2, \ldots$, we have

$$(8.28) \qquad \xi_n = \frac{b_n}{a_n + \xi_{n+1}} \quad\Longrightarrow\quad \xi_{n+1} = \frac{b_n}{\xi_n} - a_n.$$

By assumption, we have $0 < b_n \le a_n$ for all n. It follows that $\xi_n > 0$ for all n and therefore

$$\xi_n = \frac{b_n}{a_n + \xi_{n+1}} < \frac{b_n}{a_n} \le 1,$$

therefore $0 < \xi_n < 1$ for all n. Since $\xi_0 = \xi$, which is rational by assumption, by the second equality in (8.28) and induction it follows that ξ_n is rational for all n. Since $0 < \xi_n < 1$ for all n, we can therefore write $\xi_n = s_n/t_n$ where $0 < s_n < t_n$ for all n with s_n and t_n relatively prime integers. Now from the second equality in (8.28) we see that

$$\frac{s_{n+1}}{t_{n+1}} = \xi_{n+1} = \frac{b_n}{\xi_n} - a_n = \frac{b_n t_n}{s_n} - a_n = \frac{b_n t_n - a_n s_n}{s_n}.$$

Hence,

$$s_n\, s_{n+1} = (b_n t_n - a_n s_n) t_{n+1}.$$

Thus, $t_{n+1} | s_n\, s_{n+1}$. By assumption, s_{n+1} and t_{n+1} are relatively prime, so t_{n+1} must divide s_n. In particular, $t_{n+1} < s_n$. However, $s_n < t_n$ by assumption, so $t_{n+1} < t_n$. In summary, $\{t_n\}$ is a sequence of positive integers satisfying

$$t_1 > t_2 > t_3 > \cdots > t_n > t_{n+1} > \cdots > 0,$$

which of course is an absurdity because we would eventually reach zero! □

Example 8.16. (**Irrationality of** e**, Proof III**) Since we already know that (see (8.14))

$$e = 2 + \frac{2}{2+}\,\frac{3}{3+}\,\frac{4}{4+}\,\frac{5}{5+}\,\cdots,$$

we certainly have $b_n \le a_n$ for all n, hence e is irrational!

As another application of this theorem, we get

COROLLARY 8.16. *Any infinite simple continued fraction represents an irrational number. In particular, a real number is irrational if and only if it can be represented by an infinite simple continued fraction.*

Indeed, for a simple continued fraction we have $b_n = 1$ for all n, so $0 < b_n \le a_n$ for all $n \ge 1$ holds.

EXERCISES 8.4.

1. (a) Use the simple continued fraction algorithm to the find the expansions of
$$(a)\ \sqrt{2}\quad,\quad (b)\ \frac{1-\sqrt{8}}{2}\quad,\quad (c)\ \sqrt{19}\quad,\quad (d)\ 3.14159\quad,\quad (e)\ \sqrt{7}.$$
(b) Find the value of the continued fraction expansions
$$(a)\ 4+\frac{2}{8+}\,\frac{2}{8+}\,\frac{2}{8+}\cdots\quad,\quad (b)\ \langle\overline{3}\rangle = \langle 3;3,3,3,3,3,\ldots\rangle.$$
The continued fraction in (a) was studied by Pietro Antonio Cataldi (1548–1626) and is one of the earliest infinite continued fractions on record.
2. In Section 8.7, we will prove the conjectures you make in (a) and (b) below.
(a) Using a calculator, we find that $e \approx 2.718281828$. Verify that $2.718281828 = \langle 2;1,2,1,1,4,1,1,6,\ldots\rangle$. From this, conjecture a formula for a_n, $n = 0,1,2,3,\ldots$, in the canonical continued fraction expansion for e.
(b) Using a calculator, we find that $\frac{e+1}{e-1} \approx 2.1639534137$. Find a_0, a_1, a_2, a_3 in the canonical continued fraction expansion for 2.1639534137 and conjecture a formula for a_n, $n = 0,1,2,3,\ldots$, in the canonical continued fraction expansion for $\frac{e+1}{e-1}$.
3. Let $n \in \mathbb{N}$. Prove that $\sqrt{n^2+1} = \langle n;\overline{2n}\rangle$ by using the simple continued fraction algorithm on $\sqrt{n^2+1}$. Using the same technique, find the canonical expansion of $\sqrt{n^2+2}$. (See Problem 5 below for other proofs.)
4. In this problem we show that any positive real number can be written as two different infinite continued fractions. Let a be a positive real number. Prove that
$$a = 1+\cfrac{k}{1+\cfrac{k}{1+\cfrac{k}{1+\ddots}}} = \cfrac{\ell}{1+\cfrac{\ell}{1+\cfrac{\ell}{1+\ddots}}},$$
where $k = a^2 - a$ and $\ell = a^2 + a$. Suggestion: Link the limits of continued fractions on the right to the quadratic equations $x^2 - x - k = 0$ and $x^2 + x - \ell = 0$, respectively. Find neat infinite continued fractions for $1, 2$, and 3.
5. Let x be any positive real number and suppose that $x^2 - ax - b = 0$ where a, b are positive. Prove that
$$x = a + \frac{b}{a+}\,\frac{b}{a+}\,\frac{b}{a+}\,\frac{b}{a+}\,\frac{b}{a+}\cdots.$$
Using this, prove that for any $\alpha, \beta > 0$,
$$\sqrt{\alpha^2+\beta} = \alpha + \frac{\beta}{2\alpha+}\,\frac{\beta}{2\alpha+}\,\frac{\beta}{2\alpha+}\,\frac{\beta}{2\alpha+}\cdots.$$
6. (a) Prove that a continued fraction $a_0 + \frac{b_1}{a_1+}\frac{b_2}{a_2+}\frac{b_3}{a_3+}\cdots$ converges if and only if
$$c_0 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} b_1 b_2 \cdots b_n}{q_n\, q_{n-1}}$$
converges, in which case, this sum is exactly $a_0 + \frac{b_1}{a_1+}\frac{b_2}{a_2+}\frac{b_3}{a_3+}\cdots$. Suggestion: Consider the telescoping sum $c_n = c_0 + (c_1 - c_0) + (c_2 - c_1) + \cdots + (c_n - c_{n-1})$. In

particular, for a simple continued fraction $\xi = \langle a_0; a_1, a_2, a_3, \ldots\rangle$, we have

$$\xi = 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{q_n\, q_{n-1}}.$$

(b) Assume that $\xi = a_0 + \frac{b_1}{a_1 +}\frac{b_2}{a_2 +}\frac{b_3}{a_3 +}\cdots$ converges. Prove that

$$\xi = c_0 + \sum_{n=2}^{\infty} \frac{(-1)^n a_n b_1 b_2 \cdots b_{n-1}}{q_n\, q_{n-2}}.$$

In particular, for a simple continued fraction $\xi = \langle a_0; a_1, a_2, a_3, \ldots\rangle$, we have

$$\xi = 1 + \sum_{n=2}^{\infty} \frac{(-1)^n a_n}{q_n\, q_{n-2}}.$$

7. Let $\{c_n\}$ be the convergents of $\Phi = \langle 1; 1, 1, 1, 1, 1, 1, \ldots\rangle$.
   (1) Prove that for $n \geq 1$, we have $\frac{F_{n+1}}{F_n} = c_{n-1}$. (That is, $p_n = F_{n+2}$ and $q_n = F_{n+1}$.) Conclude that

   $$\boxed{\Phi = \lim_{n\to\infty} \frac{F_{n+1}}{F_n}\ ,}$$

   a beautiful (but nontrivial) fact!

   (2) Using the previous problem, prove the incredibly beautiful formulas

   $$\boxed{\Phi = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{F_n F_{n+1}} \qquad \text{and} \qquad \Phi^{-1} = \sum_{n=2}^{\infty} \frac{(-1)^n}{F_n F_{n+2}}.}$$

8. Let $\alpha = \langle a_0; a_1, a_2, \ldots\rangle$, $\beta = \langle b_0; b_1, b_2, \ldots\rangle$ be infinite simple continued fractions. Prove that if $\alpha = \beta$, then $a_k = b_k$ for all $k = 0, 1, 2, \ldots$, which shows that the canonical simple fraction expansion of an irrational real number is unique. See Problem 6 in Exercises 8.3 for the rational case.
9. A continued fraction $a_0 + \frac{1}{a_1 +}\frac{1}{a_2 +}\frac{1}{a_3 +}\frac{1}{a_4 +}\cdots$ where the a_n are real numbers with $a_n > 0$ for $n \geq 1$ is said to be **unary**. In this problem we prove that a unary continued fraction converges if and only if $\sum a_n = \infty$. Henceforth, let $a_0 + \frac{1}{a_1 +}\frac{1}{a_2 +}\frac{1}{a_3 +}\cdots$ be unary.
   (i) Prove that $q_n \leq \prod_{k=1}^{n}(1 + a_k)$.
   (ii) Using the inequality derived in (i), prove that if the unary continued fraction converges, then $\sum a_n = \infty$.
   (iii) Prove that

   $$q_{2n} \geq 1 + a_1(a_2 + a_4 + \cdots + a_{2n}) \ \ , \ \ q_{2n-1} \geq a_1 + a_3 + \cdots + a_{2n-1},$$

   where the first inequality holds for $n \geq 1$ and the second for $n \geq 2$.
   (iv) Using the inequalities derived in (9iii), prove that if $\sum a_n = \infty$, then the unary continued fraction converges.

8.5. Diophantine approximations and the mystery of π solved!

For practical purposes, it is necessary to approximate irrational numbers by rational numbers. Also, if a rational number has a very large denominator, e.g. $\frac{1234567}{1211109876543 21}$, then it is hard to work with, so for practical purposes it would be nice to have a "good" approximation to such a rational number by a rational number with a more manageable denominator. Diophantine approximations is the subject of finding "good" or even "best" rational approximations to real numbers. Continued fractions turn out to play a very important role in this subject, to which this section is devoted. We start with a journey concerning the mysterious fraction representations of π.

8.5.1. The mystery of π and good and best approximations.

Here we review some approximations to $\pi = 3.14159265\ldots$ that have been discovered throughout the centuries (see Section 4.10 for a thorough study):

(1) 3 in the Holy Bible circa 1000 B.C. by the Hebrews; See Book of I Kings, Chapter 7, verse 23, and Book of II Chronicles, Chapter 4, verse 2:

> And he made a molten sea, ten cubits from the one brim to the other:
> it was round all about, and his height was five cubits: and a line of
> thirty cubits did compass it about. I Kings 7:23.

(2) $22/7 = 3.14285714\ldots$ (correct to two decimal places) by Archimedes of Syracuse (287 B.C. –212 B.C.) circa 250 B.C.

(3) $333/106 = 3.14150943\ldots$ (correct to four decimal places), a lower bound found by Adriaan Anthoniszoon (1527–1607) circa 1600 A.D.

(4) $355/113 = 3.14159292\ldots$ (correct to six decimal places) by Tsu Chung-Chi (429–501) circa 500 A.D.

Hmmm... these numbers certainly seem familiar! These numbers are exactly the first four convergents of the continued fraction expansion of π that we worked out in Subsection 8.4.4! From this example, it seems like approximating real numbers by rational numbers is intimately related to continued fractions; this is indeed the case as we shall see. To start our adventure in approximations, we start with the concepts of "good" and "best" approximations.

A rational number p/q is called a **good approximation** to a real number ξ if[3]

$$\boxed{\text{for all rational } \frac{a}{b} \neq \frac{p}{q} \text{ with } 1 \leq b \leq q, \text{ we have } \left|\xi - \frac{p}{q}\right| < \left|\xi - \frac{a}{b}\right|;}$$

in other words, we cannot get closer to the real number ξ with a different rational number having a denominator $\leq q$.

Example 8.17. 4/1 is *not* a good approximation to π because 3/1, which has an equal denominator, is closer to π:

$$\left|\pi - \frac{3}{1}\right| = 0.141592\ldots < \left|\pi - \frac{4}{1}\right| = 0.858407\ldots.$$

Example 8.18. As another example, 7/2 is *not* a good approximation to π because 3/1, which has a smaller denominator than 7/2, is closer to π:

$$\left|\pi - \frac{3}{1}\right| = 0.141592\ldots < \left|\pi - \frac{7}{2}\right| = 0.358407\ldots.$$

This example shows that you wouldn't want to approximate π with 7/2 because you can approximate it with the "simpler" number 3/1 that has a smaller denominator.

[3] Warning: Some authors define good approximation as: $\frac{p}{q}$ is a good approximation to ξ if for all rational $\frac{a}{b}$ with $1 \leq b < q$, we have $\left|\xi - \frac{p}{q}\right| < \left|\xi - \frac{a}{b}\right|$. This definition, although only slightly different from ours, makes some proofs *considerably easier*. Moreover, with this definition, $1,000,000/1$ is a good approximation to π (why?)! (In fact, *any* integer, no matter how big, is a good approximation to π.) On the other hand, with the definition we used, the only integer that is a good approximation to π is 3. Also, some authors define best approximation as: $\frac{p}{q}$ is a best approximation to ξ if for all rational $\frac{a}{b}$ with $1 \leq b < q$, we have $|q\xi - p| < |b\xi - a|$; with this definition of "best," one can shorten the proof of Theorem 8.20 — but then one must live with the fact that $1,000,000/1$ is a best approximation to π.

Example 8.19. On the other hand, 13/4 is a good approximation to π. This is because

$$\left|\pi - \frac{13}{4}\right| = 0.108407\ldots,$$

and there are no fractions closer to π with denominator 4, and the closest distinct fractions with the smaller denominators 1, 2, and 3 are 3/1, 7/2, and 10/3, which satisfy

$$\left|\pi - \frac{3}{1}\right| = 0.141592\ldots \quad , \quad \left|\pi - \frac{7}{2}\right| = 0.358407\ldots \quad , \quad \left|\pi - \frac{10}{3}\right| = 0.191740\ldots.$$

Thus,

$$\text{for all rational } \frac{a}{b} \neq \frac{13}{4} \text{ with } 1 \leq b \leq 4, \text{ we have } \left|\pi - \frac{13}{4}\right| < \left|\pi - \frac{a}{b}\right|.$$

Now one can argue: Is 13/4 really that great of an approximation to π? For although 3/1 is not as close to π, it is certainly much easier to work with than 13/4 because of the larger denominator 4 — moreover, we have $13/4 = 3.25$, so we didn't even gain a single decimal place of accuracy in going from 3.00 to 3.25. These are definitely valid arguments. One can also see the validity of this argument by combining fractions in the inequality in the definition of good approximation: p/q is a good approximation to ξ if

$$\text{for all rational } \frac{a}{b} \neq \frac{p}{q} \text{ with } 1 \leq b \leq q, \text{ we have } \frac{|q\xi - p|}{q} < \frac{|b\xi - a|}{b},$$

where we used that $q, b > 0$. Here, we can see that $\frac{|q\xi - p|}{q} < \frac{|b\xi - a|}{b}$ may hold not because p/q is dramatically much closer to ξ than is a/b but simply because q is a lot larger than b (like in the case 13/4 and 3/1 where 4 is much larger than 1). To try and correct this somewhat misleading notion of "good" we introduce the concept of a "best" approximation by clearing the denominators.

A rational number p/q is called a **best approximation** to a real number ξ if

$$\boxed{\text{for all rational } \frac{a}{b} \neq \frac{p}{q} \text{ with } 1 \leq b \leq q, \text{ we have } \left|q\xi - p\right| < \left|b\xi - a\right|.}$$

Example 8.20. We can see that $p/q = 13/4$ is *not* a best approximation to π because with $a/b = 3/1$, we have $1 \leq 1 \leq 4$ yet

$$\left|4 \cdot \pi - 13\right| = 0.433629\ldots \not< \left|1 \cdot \pi - 3\right| = 0.141592\ldots.$$

Thus, 13/4 is a good approximation to π but is far from a best approximation.

In the following proposition, we show that any best approximation is a good one.

Proposition 8.17. *A best approximation is a good one, but not vice versa.*

Proof. We already gave an example showing that a good approximation may not be a best one, so let p/q be a best approximation to ξ; we shall prove that p/q is a good one too. Let $a/b \neq p/q$ be rational with $1 \leq b \leq q$. Then $|q\xi - p| < |b - \xi a|$ since p/q is a best approximation, and also, $\frac{1}{q} \leq \frac{1}{b}$ since $b \leq q$, hence

$$\left|\xi - \frac{p}{q}\right| = \frac{|q\xi - p|}{q} < \frac{|b\xi - a|}{q} \leq \frac{|b\xi - a|}{b} = \left|\xi - \frac{a}{b}\right| \quad \Longrightarrow \quad \left|\xi - \frac{p}{q}\right| < \left|\xi - \frac{a}{b}\right|.$$

This shows that p/q is a good approximation. $\square$

In the following subsection, we shall prove the best approximation theorem, Theorem 8.20, which states that

> *(**Best approximation theorem**) Every best approximation of a real number (rational or irrational) is a convergent of its canonical continued fraction expansion and conversely, each of the convergents $c_1, c_2, c_3, \ldots$ is a best approximation.*

8.5.2. Approximations, convergents, and the "most irrational" of all irrational numbers.

The objective of this subsection is to understand how convergents approximate real numbers. In the following theorem, we show that the convergents of the simple continued fraction of a real number ξ get increasingly closer to ξ. (See Problem 4 for the general case of nonsimple continued fractions.)

Theorem 8.18 (**Fundamental approximation theorem**). *Let ξ be an irrational number and let $\{c_n = p_n/q_n\}$ be the convergents of its canonical continued fraction. Then the following inequalities hold:*

$$\bigl|\xi - c_n\bigr| < \frac{1}{q_n q_{n+1}}, \quad \bigl|\xi - c_{n+1}\bigr| < \bigl|\xi - c_n\bigr|, \quad \bigl|q_{n+1}\xi - p_{n+1}\bigr| < \bigl|q_n\xi - p_n\bigr|.$$

If ξ is a rational number and the convergent c_{n+1} is defined (that is, if $\xi \neq c_n$), then these inequalities still hold, with the exception that $|\xi - c_n| = \frac{1}{q_n q_{n+1}}$ if $\xi = c_{n+1}$.

Proof. We prove this theorem for ξ irrational; the rational case is proved using a similar argument, which we leave to you if you're interested. The proof of this theorem is very simple. We just need the inequalities (see Corollary 8.13)

$$c_n < c_{n+2} < \xi < c_{n+1} \qquad \text{or} \qquad c_{n+1} < \xi < c_{n+2} < c_n, \tag{8.29}$$

depending on whether n is even or odd, respectively, and the fundamental recurrence relations (see Corollary 8.6):

$$c_{n+1} - c_n = \frac{(-1)^n}{q_n\, q_{n+1}} \quad , \quad c_{n+2} - c_n = \frac{(-1)^n a_{n+2}}{q_n\, q_{n+2}}. \tag{8.30}$$

Now the first inequality of our theorem follows easily:

$$\bigl|\xi - c_n\bigr| \overset{\text{by (8.29)}}{<} \bigl|c_{n+1} - c_n\bigr| \overset{\text{by (8.30)}}{=} \left|\frac{(-1)^n}{q_n\, q_{n+1}}\right| = \frac{1}{q_n\, q_{n+1}}.$$

We now prove that $\bigl|q_{n+1}\xi - p_{n+1}\bigr| < \bigl|q_n\xi - p_n\bigr|$. To prove this, we work on the left and right-hand sides separately. For the left-hand side, we have

$$\begin{aligned}\bigl|q_{n+1}\xi - p_{n+1}\bigr| = q_{n+1}\left|\xi - \frac{p_{n+1}}{q_{n+1}}\right| = q_{n+1}\bigl|\xi - c_{n+1}\bigr| &< q_{n+1}\bigl|c_{n+2} - c_{n+1}\bigr| \text{ by (8.29)}\\ &= q_{n+1}\frac{1}{q_{n+1}\, q_{n+2}} \quad \text{by (8.30)}\\ &= \frac{1}{q_{n+2}}.\end{aligned}$$

Hence, $\frac{1}{q_{n+2}} > \big|q_{n+1}\xi - p_{n+1}\big|$. Now,

$$\begin{aligned}\big|q_n\xi - p_n\big| = q_n\Big|\xi - \frac{p_n}{q_n}\Big| = q_n\big|\xi - c_n\big| &> q_n\big|c_{n+2} - c_n\big| \qquad \text{by (8.29)}\\ &= q_n\frac{a_{n+2}}{q_n\, q_{n+2}} \qquad \text{by (8.30)}\\ &= \frac{a_{n+2}}{q_{n+2}} \geq \frac{1}{q_{n+2}} > \big|q_{n+1}\xi - p_{n+1}\big|.\end{aligned}$$

This proves our third inequality. Finally, using what we just proved, and that

$$q_{n+1} = a_{n+1}q_n + q_{n-1} \geq q_n + q_{n-1} > q_n \quad \Longrightarrow \quad \frac{1}{q_{n+1}} < \frac{1}{q_n},$$

we see that

$$\begin{aligned}\big|\xi - c_{n+1}\big| = \Big|\xi - \frac{p_{n+1}}{q_{n+1}}\Big| &= \frac{1}{q_{n+1}}\big|q_{n+1}\xi - p_{n+1}\big|\\ &< \frac{1}{q_{n+1}}\big|q_n\xi - p_n\big|\\ &< \frac{1}{q_n}\big|q_n\xi - p_n\big| = \Big|\xi - \frac{p_n}{q_n}\Big| = \big|\xi - c_n\big|.\end{aligned}$$

□

It is important to only use the *canonical* expansion when ξ is rational. This is because the statement that $\big|q_{n+1}\xi - p_{n+1}\big| < \big|q_n\xi - p_n\big|$ may not *not* be true if we don't use the canonical expansion.

Example 8.21. Consider 5/3, which has the canonical expansion:

$$\frac{5}{3} = \langle 1; 1, 2\rangle = 1 + \cfrac{1}{1 + \cfrac{1}{2}}.$$

We can write this as the noncanonical expansion by breaking up the 2:

$$\xi = \langle 1; 1, 1, 1\rangle = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1}}} = \frac{5}{3}.$$

The convergents for this noncanonical expansion of ξ are $c_0 = 1/1$, $c_2 = 2/1$, $c_3 = 3/2$, and $\xi = c_4 = 5/3$. In this case,

$$\big|q_3\xi - p_3\big| = \Big|2\cdot\frac{5}{3} - 3\Big| = \frac{1}{3} = \Big|1\cdot\frac{5}{3} - 2\Big| = \big|q_2\xi - p_2\big|,$$

so for this example, $\big|q_2\xi - p_2\big| \not< \big|q_2\xi - p_2\big|$.

We now discuss the "most irrational" of all irrational numbers. From the best approximation theorem (Theorem 8.20 we'll prove in a moment) we know that the best approximations of a real number ξ are convergents and from the fundamental approximation theorem 8.18, we have the error estimate

$$\big|\xi - c_n\big| < \frac{1}{q_nq_{n+1}} \quad \Longrightarrow \quad \big|q_n\xi - p_n\big| < \frac{1}{q_{n+1}}. \tag{8.31}$$

This shows you that the larger the q_n's are, the better the best approximations are. Since the q_n's are determined by the recurrence relation $q_n = a_n q_{n-1} + q_{n-2}$, we see that the larger the a_n's are, the larger the q_n's are. In summary, ξ can be approximated very "good" by rational numbers when it has *large* a_n's and very "bad" by rational numbers when it has *small* a_n's.

Example 8.22. Here is a "good" example: Recall from (8.27) the continued fraction for π: $\pi = \langle 3; 7, 15, 1, 292, 1, 1, 1, 2, 1, \ldots \rangle$, which has convergents $c_0 = 3$, $c_1 = \frac{22}{7}$, $c_2 = \frac{333}{106}$, $c_3 = \frac{355}{113}$, $c_4 = \frac{103993}{33102}$, ... Because of the large number $a_4 = 292$, we see from (8.31) that we can approximate π very nicely with c_3: Using the left-hand equation in (8.31), we see that

$$\left|\pi - c_3\right| < \frac{1}{q_3 q_4} = \frac{1}{113 \cdot 33102} = 0.000000267\ldots,$$

which implies that $c_3 = \frac{355}{113}$ approximates π to within six decimal places! (Just to check, note that $\pi = 3.14159265\ldots$ and $\frac{355}{113} = 3.14159292\ldots$.) It's amazing how many decimal places of accuracy we can get with just taking the c_3 convergent!

Example 8.23. (**The "most irrational" number**) Here is a "bad" example: From our discussion after (8.31), we saw that the smaller the a_n's are, the worse it can be approximated by rationals. Of course, since 1 is the smallest natural number, we can consider the golden ratio

$$\Phi = \frac{1+\sqrt{5}}{2} = \langle 1; 1, 1, 1, 1, 1, 1, 1, \ldots \rangle = 1.6180339887\ldots$$

as being the "worst" of all irrational numbers that can be approximated by rational numbers. Indeed, we saw that we could get six decimal places of π by just taking c_3; for Φ we need c_{18}! (Just to check, we find that $c_{17} = \frac{4181}{2584} = 1.6180340557\ldots$ — not quite six decimals — and $c_{18} = \frac{6765}{4181} = 1.618033963\ldots$ — got the sixth one. Also notice the large denominator 4181 just to get six decimals.) Therefore, Φ wins the prize for the "most irrational" number in that it's the "farthest" from the rationals! We continue our discussion on "most irrational" in Subsection 8.10.3.

We now show that best approximations are exactly convergents; this is one of the most important properties of continued fractions. We first need the following lemma, whose ingenious proof we learned from Beskin's beautiful book [**28**].

Lemma 8.19. *If p_n/q_n, $n \geq 0$, is a convergent of the canonical continued fraction expansion of a real number ξ and $p/q \neq p_n/q_n$ is a rational number with $q > 0$ and $1 \leq q < q_{n+1}$, then*

$$|q_n \xi - p_n| \leq |q\xi - p|.$$

Moreover, this inequality is an equality if and only if

$$\xi = \frac{p_{n+1}}{q_{n+1}} \ , \quad p = p_{n+1} - p_n \ , \quad \text{and} \ \ q = q_{n+1} - q_n.$$

If this is the case, then we have $q > q_n$ if $n \geq 1$.

Proof. Let p_n/q_n, $n \geq 0$, be a convergent of the canonical continued fraction expansion of a real number ξ and let $p/q \neq p_n/q_n$ be a rational number with $q > 0$ and $1 \leq q < q_{n+1}$. Note that if ξ happens to be rational, we are implicity assuming that $\xi \neq p_n/q_n$ so that q_{n+1} is defined.

Step 1: The trick. To prove that $|q_n\xi - p_n| \le |q\xi - p|$, the trick is to write p and q as linear combinations of $p_n, p_{n+1}, q_n, q_{n+1}$:

$$\begin{aligned} p &= p_n x + p_{n+1} y \\ q &= q_n x + q_{n+1} y. \end{aligned} \tag{8.32}$$

Using basic linear algebra, together with the fact that $p_{n+1}q_n - p_n q_{n+1} = (-1)^n$, we can solve these simultaneous linear equations for x and y obtaining

$$x = (-1)^n\big(p_{n+1}q - pq_{n+1}\big) \quad , \quad y = (-1)^n\big(pq_n - p_n q\big).$$

These formulas are not needed below except for the important fact that these formulas show that x and y are *integers*. Now, using the formulas in (8.32), we see that

$$\begin{aligned} q\xi - p &= \big(q_n x + q_{n+1}y\big)\xi - p_n x - p_{n+1}y \\ &= \big(q_n\xi - p_n\big)x + \big(q_{n+1}\xi - p_{n+1}\big)y. \end{aligned}$$

Therefore,

$$|q\xi - p| = \big|\big(q_n\xi - p_n\big)x + \big(q_{n+1}\xi - p_{n+1}\big)y\big|. \tag{8.33}$$

Step 2: Our goal is to simplify the right-hand side of (8.33) by understanding the signs of the terms in the absolute values. First of all, since $q, q_n, q_{n+1} > 0$, from the second formula in (8.32), we see that $x \le 0$ and $y \le 0$ is not possible (for then $q \le 0$, contradicting that $q > 0$). If $x > 0$ and $y > 0$, then we would have

$$q = q_n x + q_{n+1}y > q_{n+1},$$

contradicting that $q < q_{n+1}$. (Note that $y > 0$ is the same thing as saying $y \ge 1$ because y is an integer.) If $x = 0$, then the formulas (8.32) show that $p = p_{n+1}y$ and $q = q_{n+1}y$. Since q and q_{n+1} are positive, we must have $y > 0$ and we have $q \ge q_{n+1}$, contradicting that $q < q_{n+1}$. If $y = 0$, then the formulas (8.32) show that $p = p_n x$ and $q = q_n x$, so $p/q = p_n/q_n$ and this contradicts the assumption that $p/q \ne p_n/q_n$. Summarizing our findings: We may assume that x and y are both nonzero and have opposite signs. By Corollary 8.11 we know that

$$\xi - \frac{p_n}{q_n} \quad \text{and} \quad \xi - \frac{p_{n+1}}{q_{n+1}}$$

have opposite signs. Therefore, $q_n\xi - p_n$ and $q_{n+1}\xi - p_{n+1}$ have opposite signs and hence, since x and y also have opposite signs,

$$\big(q_n\xi - p_n\big)x \quad \text{and} \quad \big(q_{n+1}\xi - p_{n+1}\big)y$$

have the same sign. Therefore, in (8.33), we have

$$\big|q\xi - p\big| = |q_n\xi - p_n|\,|x| + |q_{n+1}\xi - p_{n+1}|\,|y|.$$

Step 3: We now prove our result. Since $x \ne 0$, we have $|x| \ge 1$ (because x is an integer), so

$$|q_n\xi - p_n| \le |q_n\xi - p_n|\,|x| \le |q_n\xi - p_n|\,|x| + |q_{n+1}\xi - p_{n+1}|\,|y| = \big|q\xi - p\big|,$$

and we have proved that $|q_n\xi - p_n| \le |q\xi - p|$ just as we set out to do.

Now assume that we have an equality: $|q_n\xi - p_n| = |q\xi - p|$. Then we have

$$\begin{aligned} &|q_n\xi - p_n| = |q_n\xi - p_n|\,|x| + |q_{n+1}\xi - p_{n+1}|\,|y| \\ &\qquad \Longrightarrow \quad |q_n\xi - p_n|\,(|x| - 1) + |q_{n+1}\xi - p_{n+1}|\,|y| = 0 \end{aligned}$$

Since x and y are both nonzero integers, we have in particular, $|x| - 1 \geq 0$ and $|y| > 0$. Therefore,

$$|q_n\xi - p_n|\,(|x| - 1) + |q_{n+1}\xi - p_{n+1}|\,|y| = 0 \quad \Longleftrightarrow \quad \xi = \frac{p_{n+1}}{q_{n+1}} \text{ and } |x| = 1.$$

If $x = +1$, then $y < 0$ (because x and y have opposite signs) so $y \leq -1$ since y is an integer and hence by the second equation in (8.32), we have

$$q = q_n x + q_{n+1}y = q_n + q_{n+1}y \leq q_n - q_{n+1} \leq 0,$$

because $q_n \leq q_{n+1}$. This is impossible since $q > 0$ by assumption. Hence, $x = -1$. In this case $y > 0$ and hence $y \geq 1$. If $y \geq 2$, then

$$q = q_n x + q_{n+1}y = -q_n + q_{n+1}y \geq -q_n + 2q_{n+1} = q_{n+1} + (q_{n+1} - q_n) \geq q_{n+1},$$

which contradicts the fact that $q < q_{n+1}$. Therefore, $y = 1$. In conclusion, we have seen that $|q_n\xi - p_n| = |q\xi - p|$ if and only if $\xi = p_{n+1}/q_{n+1}$, $x = -1$ and $y = 1$, which by the formulas in (8.32), imply that $p = p_{n+1} - p_n$ and $q = q_{n+1} - q_n$. Finally, by the Wallis-Euler recurrence relations, we have

$$q = q_{n+1} - q_n = a_{n+1}q_n + q_{n-1} - q_n = (a_{n+1} - 1)q_n + q_{n-1}.$$

If $n \geq 1$, then $q_{n-1} \geq 1$ and $a_{n+1} \geq 2$. Therefore, if $n \geq 1$, then $q > q_n$ and our proof is complete. □

As an easy consequence of this lemma, it follows that every convergent p_n/q_n with $n \geq 1$ of the canonical continued fraction expansion of a real number ξ must be a best approximation. Indeed, if $\xi = p_n/q_n$, then automatically p_n/q_n is a best approximation of ξ. So assume that $\xi \neq p_n/q_n$, where $n \geq 1$, and let $p/q \neq p_n/q_n$ with $1 \leq q \leq q_n$. Then, since $n \geq 1$, we have $q_n < q_{n+1}$, so $1 \leq q < q_{n+1}$. Therefore by Lemma 8.19,

$$|q_n\xi - p_n| < |q\xi - p|,$$

since the exceptional case is ruled out ($q \not> q_n$ because $q \leq q_n$ by assumption).

Note that we left out p_0/q_0 may not be a best approximation!

Example 8.24. Consider $\sqrt{3} = 1.73205080\ldots$. The best integer approximation to $\sqrt{3}$ is 2. In Subsection 8.4.3 we found that $\sqrt{3} = \langle 1; \overline{1,2}\rangle$. Thus, $p_0/q_0 = 1$, which is not a best approximation. However, $p_1/q_1 = 1 + \frac{1}{1} = 2$ is a best approximation.

THEOREM 8.20 (**Best approximation theorem**). *Every best approximation of a real number (rational or irrational) is a convergent of its canonical continued fraction expansion and conversely, each of the convergents $c_1, c_2, c_3, \ldots$ is a best approximation.*

PROOF. We just have to prove that if p/q with $q > 0$ is a best approximation to $\xi \in \mathbb{R}$, then p/q is a convergent. Assume first that ξ is irrational. Then $1 = q_0 \leq q_1 < q_2 < \cdots < q_n \to \infty$, so we can choose a k such that

$$q_k \leq q < q_{k+1}.$$

By Lemma 8.19, if $p/q \neq p_k/q_k$, we have

$$|b\xi - a| \leq |q\xi - p|,$$

where $a = p_k$ and $b = q_k \leq q$ contradicting that p/q is a best approximation to p/q. Therefore, $p/q = p_k/q_k$, so p/q is a convergent of ξ.

Assume now that ξ is rational. Then $\xi = p_{n+1}/q_{n+1}$ for some $n = -1, 0, 1, \ldots$. We consider three cases:

Case 1: $q = q_{n+1}$: Then the assumption that p/q is a best approximation to ξ implies that $p/q = p_{n+1}/q_{n+1}$ (why?) so p/q is a convergent.

Case 2: $q > q_{n+1}$: In fact, this case cannot occur because

$$|b\xi - a| = 0 \leq |p\xi - q|$$

would hold for $a = p_{n+1}$ and $b = q_{n+1} < q$ contradicting that p/q is a best approximation to ξ.

Case 3: $1 \leq q < q_{n+1}$: Since $1 = q_0 \leq q_1 < q_2 < \cdots < q_{n+1}$ it follows that there is a k such that

$$q_k \leq q < q_{k+1}.$$

Then by Lemma 8.19, if $p/q \neq p_k/q_k$, we have

$$|b\xi - a| \leq |q\xi - p|.$$

where $a = p_k$ and $b = q_k \leq q$ contradicting that p/q is a best approximation to p/q. Therefore, $p/q = p_k/q_k$, so p/q is a convergent of ξ. □

8.5.3. Dirichlet's approximation theorem. Using Theorem 8.20, we prove the following famous fact.

THEOREM 8.21 (**Dirichlet's approximation theorem**). *Amongst two consecutive convergents $p_n/q_n, p_{n+1}/q_{n+1}$ with $n \geq 0$ of the canonical continued fraction expansion to a real number (rational or irrational) ξ, one of them satisfies*

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{2q^2}. \tag{8.34}$$

Conversely, if a rational number p/q satisfies (8.34)*, then it is a convergent.*

PROOF. We begin by proving that a rational number satisfying (8.34) must be a convergent, then we show that convergents satisfy (8.34).

Step 1: Assume that p/q satisfies (8.34). To prove that it must be a convergent, we just need to show that it is a best approximation. To this end, assume that $a/b \neq p/q$ with $b > 0$ and that

$$\big|b\xi - a\big| \leq \big|q\xi - p\big|;$$

we must show that $q < b$. To prove this, we note that (8.34) implies that

$$\left|\xi - \frac{a}{b}\right| = \frac{1}{b}\big|b\xi - a\big| \leq \frac{1}{b}\big|q\xi - p\big| < \frac{1}{b} \cdot \frac{1}{2q} = \frac{1}{2bq}.$$

This inequality plus (8.34) give

$$\left|\frac{aq - bp}{bq}\right| = \left|\frac{a}{b} - \frac{p}{q}\right| = \left|\frac{a}{b} - \xi + \xi - \frac{p}{q}\right| \leq \left|\frac{a}{b} - \xi\right| + \left|\xi - \frac{p}{q}\right| < \frac{1}{2bq} + \frac{1}{2q^2}.$$

Since $a/b \neq p/q$, $|aq - bp|$ is a positive integer, that is, $1 \leq |aq - bp|$, therefore

$$\frac{1}{bq} < \frac{1}{2bq} + \frac{1}{2q^2} \quad \Longrightarrow \quad \frac{1}{2bq} < \frac{1}{2q^2} \quad \Longrightarrow \quad \frac{1}{b} < \frac{1}{q} \quad \Longrightarrow \quad q < b,$$

just as we wanted to show. We now show that one of two consecutive convergents satisfies (8.34). Let p_n/q_n and p_{n+1}/q_{n+1}, $n \geq 0$, be two consecutive convergents.

Step 2: Assume first that $q_n = q_{n+1}$. Since $q_{n+1} = a_{n+1}q_n + q_{n-1}$ we see that $q_n = q_{n+1}$ if and only if $n = 0$ (because $q_{n-1} = 0$ if and only if $n = 0$) and

$a_1 = 1$, in which case, $q_1 = q_0 = 1$, $p_0 = a_0$, and $p_1 = a_0a_1 + 1 = a_0 + 1$. Therefore, $p_0/q_0 = a_0/1$ and $p_1/q_1 = (a_0 + 1)/1$, so we just have to show that

$$\left|\xi - a_0\right| < \frac{1}{2} \quad \text{or} \quad \left|\xi - (a_0 + 1)\right| < \frac{1}{2}.$$

But one of these must hold because $a_0 = \lfloor \xi \rfloor$, so

$$a_0 \leq \xi < a_0 + 1.$$

Note that the special situation where ξ is exactly half-way between a_0 and $a_0 + 1$, that is, $\xi = a_0 + 1/2 = \langle a_0, 2\rangle$, is not possible under our current assumptions because in this special situation, $q_1 = 2 \neq 1 = q_0$.

Step 3: Assume now that $q_n \neq q_{n+1}$. Consider two consecutive convergents c_n and c_{n+1}. We know that either

$$c_n < \xi < c_{n+1} \quad \text{or} \quad c_{n+1} < \xi < c_n,$$

depending on whether n is even or odd. For concreteness, assume that n is even; the odd case is entirely similar. Then from $c_n < \xi < c_{n+1}$ and the fundamental recurrence relation $c_{n+1} - c_n = 1/q_nq_{n+1}$, we see that

$$\left|\xi - c_n\right| + \left|c_{n+1} - \xi\right| = (\xi - c_n) + (c_{n+1} - \xi) = c_{n+1} - c_n = \frac{1}{q_nq_{n+1}}.$$

Now observe that since $q_n \neq q_{n+1}$, we have

$$0 < \frac{1}{2}\Big(\frac{1}{q_n} - \frac{1}{q_{n+1}}\Big)^2 = \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2} - \frac{1}{q_n\, q_{n+1}} \quad \Longrightarrow \quad \frac{1}{q_nq_{n+1}} < \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2},$$

so

$$\left|\xi - c_n\right| + \left|\xi - c_{n+1}\right| < \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2}. \tag{8.35}$$

It follows that $|\xi - c_n| < 1/2q_n^2$ or $\left|\xi - c_{n+1}\right| < 1/2q_{n+1}^2$, otherwise (8.35) would fail to hold. This completes our proof. □

Exercises 8.5.

1. In this problem we find all the good approximations to $2/7$. First, to see things better, let's write down the some fractions with denominators less than 7:
$$\frac{0}{1} < \frac{1}{6} < \frac{1}{5} < \frac{1}{4} < \frac{2}{7} < \frac{1}{3} < \frac{2}{5} < \frac{1}{2}.$$
By examining the absolute values $\left|\xi - \frac{a}{b}\right|$ for the fractions listed, show that the good approximations to $2/7$ are $0/1, 1/2, 1/3, 1/4$, and of course, $2/7$. Now let's find which of the good approximations are best *without* using the best approximation theorem. To do so, compute the absolute values
$$\left|1 \cdot \frac{2}{7} - 0\right| \ , \ \left|2 \cdot \frac{2}{7} - 1\right| \ , \ \left|3 \cdot \frac{2}{7} - 1\right| \ , \ \left|4 \cdot \frac{2}{7} - 1\right|$$
and from these numbers, determine which of the good approximations are best. Using a similar method, find the good and best approximations to $3/7$, $3/5$, $8/5$, and $2/9$.
2. Prove that a real number ξ is irrational if and only if there are infinitely many rational numbers p/q satisfying
$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^2}.$$
3. In this problem we find very beautiful approximations to π.

(a) Using the canonical continued fraction algorithm, prove that

$$\pi^4 = 97.40909103400242\ldots = \langle 97, 2, 2, 3, 1, 16539, 1, \ldots\rangle.$$

(Warning: If your calculator doesn't have enough decimal places of accuracy, you'll probably get a different value for 16539.)

(b) Compute $c_4 = \frac{2143}{22}$ and therefore, $\pi \approx \left(\frac{2143}{22}\right)^{1/4}$. Note that $\pi = 3.141592653\ldots$ while $(2143/22)^{1/4} = 3.141592652$, quite accurate! This approximation is due to Srinivasa Aiyangar Ramanujan (1887–1920) [**27**, p. 160].[4] As explained on Weinstein's website [**240**], we can write this approximation in **pandigital** form, that is, using all digits $0, 1, \ldots, 9$ exactly once :

$$\boxed{\pi \approx \left(\frac{2143}{22}\right)^{1/4} = \sqrt{\sqrt{0 + 3^4 + \frac{19^2}{78-56}}}.}$$

(c) By determining certain convergents of the continued fraction expansions of π^2, π^3, and π^5, derive the equally fascinating results:

$$\boxed{\pi \approx \sqrt{10}\ , \quad \left(\frac{227}{23}\right)^{1/2}\ , \quad 31^{1/3}\ , \quad \left(\frac{4930}{159}\right)^{1/3}\ , \quad 306^{1/5}\ , \quad \left(\frac{77729}{254}\right)^{1/5}\ .}$$

The approximation $\pi \approx \sqrt{10} = 3.162\ldots$ was known in Mesopotamia thousands of years before Christ [**170**]!

4. If $c_n = a_0 + \frac{b_1}{a_1} + \cdots + \frac{b_n}{a_n}$ and $\xi = a_0 + \frac{b_1}{a_1} + \frac{b_2}{a_2} + \cdots$, where $a_n \geq 1$ for $n \geq 1$, $b_n > 0$, and $\sum_{n=1}^{\infty} \frac{a_n a_{n+1}}{b_{n+1}} = \infty$, prove that for any $n = 0, 1, 2, \ldots$, we have $|\xi - c_{n+1}| < |\xi - c_n|$ and $|q_{n+1}\xi - p_{n+1}| < |q_n\xi - p_n|$ (cf. Theorem 8.18).

5. (**Pythagorean triples**) Please review Problems 8 and 9 in Exercises 2.4 concerning primitive Pythagorean triples. Following Schillo [**201**] we ask the following question: Given a right triangle, is there a primitive right triangle similar to it? The answer is "not always" since e.g. the triangle with sides $(1, 1, \sqrt{2})$ is not similar to any triangle with integer sides (why?). So we ask: Given a right triangle, is there a primitive right triangle "nearly" similar to it? The answer is "yes" and here's one way to do it.

(i) Given a right triangle $\triangle$, let θ be one of its acute angles. Prove that if $\tan(\theta/2) = p/q$ where $p, q \in \mathbb{Z}$ have no common factors with $q > 0$, then $\tan\theta = 2pq/(q^2 - p^2)$. Furthermore, prove that if (x, y, z) is a Pythagorean triple where $x/y = 2pq/(q^2 - p^2)$, then (x, y, z) is similar to $\triangle$. Of course, in general $\tan(\theta/2)$ is not rational so $\tan(\theta/2) = p/q$ cannot hold. However, if $\tan(\theta/2) \approx p/q$, we have $\tan\theta \approx 2pq/(q^2 - p^2)$, so if there exists a Pythagorean triple (x, y, z) where $x/y = 2pq/(q^2 - p^2)$, then (x, y, z) is nearly similar to $\triangle$. Now by Problem 9 in Exercises 2.4, there does exist such a triple:

$$(8.36)\qquad (x, y, z) \text{ is primitive,} \quad \text{where } \begin{cases} x = 2pq\,,\ y = q^2 - p^2\,,\ z = p^2 + q^2, & \text{or,} \\ x = pq\,,\ y = \frac{q^2 - p^2}{2}\,,\ z = \frac{p^2+q^2}{2}, \end{cases}$$

according as p and q have opposite or the same parity. Notice that in either case, we have $x/y = 2pq/(q^2 - p^2)$. In summary, to find a primitive right triangle "nearly" similar to our given triangle, we just have to approximate $\tan(\theta/2)$ by rational numbers and let (x, y, z) be given by (8.36) ... approximating $\tan(\theta/2)$ by rational numbers is where continued fractions come in!

(ii) Let's apply Part (i) to the triangle $(1, 1, \sqrt{2})$. In this case, $\theta = 45°$. Prove that $\tan(\theta/2) = \sqrt{2} - 1$. Prove that the convergents of the continued fraction expansion of $\sqrt{2} - 1$ are of the form $c_n = u_n/u_{n+1}$ where $u_n = 2u_{n-1} + u_{n-2}$ $(n \geq 2)$ with

[4] *An equation means nothing to me unless it expresses a thought of God. Srinivasa Ramanujan (*1887–1920*).*

$u_0 = 0$, $u_1 = 1$. Prove that (x_n, y_n, z_n), where $x_n = 2u_n u_{n+1}$, $y_n = u_{n+1}^2 - u_n^2$, and $z_n = u_{n+1}^2 + u_n^2$, forms a sequence of primitive Pythagorean triples. (This sequence gives triangles that are more and more similar to $(1, 1, \sqrt{2})$ as $n \to \infty$.)

8.6. ★ Continued fractions and calendars, and math and music

We now do some fun stuff with continued fractions and their applications to calendars and pianos! In the exercises, you'll see how Christian Huygens (1629–1695), a Dutch physicist, made his model of the solar system (cf. [**147**]).

8.6.1. Calendars. Calendar making is an amazing subject; see Tøndering's (free!) book [**224**] for a fascinating look at calendars. A year, technically a **tropical year**, is the time it takes from one vernal equinox to the next. Recall that there are two equinoxes, which is basically (there is a more technical definition) the time when night and day have the same length. The vernal equinox occurs around March 21, the first day of spring, and the autumnal equinox occurs around September 23, the first day of fall. A year is approximately 365.24219 days. As you might guess, not being a whole number of days makes it quite difficult to make accurate calenders, and for this reason, the art of calendar making has been around since the beginning. Here are some approximations to a year that you might know about:

(1) 365 days, the ancient Egyptians and others.
(2) $365\frac{1}{4}$ days, Julius Caesar (100 B.C.–44 B.C.), 46 B.C., giving rise to the **Julian calendar**.
(3) $365\frac{97}{400}$ days, Pope Gregory XIII (1502–1585), 1585, giving rise to the **Gregorian calendar**, the calendar that is now the most widely-used calendar.

See Problem 1 for Persian calenders and their link to continued fractions. Let us analyze these calenders more thoroughly. First, the ancient calendar consisting of 365 days. Since a true year is approximately 365.24219 days, an ancient year has

$$0.24219 \textit{ less} \text{ days than a true year.}$$

Thus, after 4 years, with an ancient calendar you'll lose approximately

$$4 \times .24219 = 0.9687 \text{ days } \approx 1 \text{ day.}$$

After 125 years, with an ancient calendar you'll lose approximately

$$125 \times .24219 = 30.27375 \text{ days } \approx 1 \text{ month.}$$

So, instead of having spring around March 21, you'll have it in February! After 500 years, with an ancient calendar you'll lose approximately

$$500 \times .24219 = 121.095 \text{ days } \approx 4 \text{ months.}$$

So, instead of having spring around March 21, you'll have it in November! As you can see, this is getting quite ridiculous.

In the Julian calendar, there are an average of $365\frac{1}{4}$ days in a Julian year. The fraction $\frac{1}{4}$ is played out as we all know: We add *one* day to the ancient calendar every *four* years giving us a "leap year", that is, a year with 366 days. Thus, just as we said, a Julian calendar year gives the estimate

$$\frac{4 \times 365 + 1 \text{ days}}{4 \text{ years}} = 365\frac{1}{4} \frac{\text{days}}{\text{year}}.$$

The Julian year has

$$365.25 - 365.24219 = 0.00781 \textit{ more} \text{ days than a true year.}$$

So, for instance, after 125 years, with a Julian calendar you'll gain

$$125 \times .00781 = 0.97625 \text{ days } \approx 1 \text{ day.}$$

Not bad. After 500 years, with a Julian calendar you'll gain

$$500 \times .00781 = 3.905 \text{ days } \approx 4 \text{ days.}$$

Again, not bad! But, still, four days gained is still four days gained.

In the Gregorian calendar, there are an average of $365\frac{97}{400}$ days, that is, we add *ninety seven* days to the ancient calendar every *four hundred* years. These extra days are added as follows: Every four years we add one extra day, a "leap year" just like in the Julian calendar — however, this gives us 100 extra days in 400 years; so to offset this, we do not have a leap year for the century marks except 400, 800, 1200, 1600, 2000, 2400, ... multiples of 400. For example, consider the years

$$1604, 1608, \ldots, 1696, 1700, 1704, \ldots, 1796, 1800, 1804, \ldots, 1896,$$
$$1900, 1904, \ldots, 1996, 2000.$$

Each of these years is a leap year except the three years 1700, 1800, and 1900 (but note that the year 2000 was a leap year since it is a multiple of 400, as you can verify on your old calendar). Hence, in the four hundred years from the end of 1600 to the end of 2000, we added only 97 total days since we didn't add extra days in 1700, 1800, and 1900. So, just as we said, a Gregorian calendar gives the estimate

$$\frac{400 \times 365 + 97}{400} = 365\frac{97}{400}\ \frac{\text{days}}{\text{year}}.$$

Since $365\frac{97}{400} = 365.2425$, the Gregorian year has

$$365.2425 - 365.24219 = 0.00031 \textit{ more} \text{ days than a true year.}$$

For instance, after 500 years, with a Gregorian calendar you'll gain

$$500 \times 0.00031 = 0.155 \text{ days } \approx 0 \text{ days!}$$

Now let's link calendars with continued fractions. Here is the continued fraction expansion of the tropical year:

$$365.24219 = \langle 365; 4, 7, 1, 3, 24, 6, 2, 2\rangle.$$

This has convergents:

$$c_0 = 365\ ,\ c_1 = 365\frac{1}{4}\ ,\ c_2 = 365\frac{7}{29}\ ,\ c_3 = 365\frac{8}{33}\ ,\ c_4 = 365\frac{31}{128}\ , \ldots.$$

Here, we see that c_0 is the ancient calendar and c_1 is the Julian calendar, but where is the Gregorian calendar? It's not on this list, but it's almost c_3 since

$$\frac{8}{33} = \frac{8}{33} \cdot \frac{12}{12} = \frac{96}{396} \approx \frac{97}{400}.$$

However, it turns out that $c_3 = 365\frac{8}{33}$ is *exactly* the average number of days in the Persian calendar introduced by the mathematician, astronomer, and poet Omar Khayyam (1048 –1131)! See Problem 1 for the modern Persian calendar!

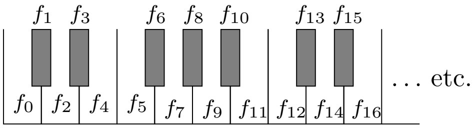


FIGURE 8.1. The k-th key, starting from $k = 0$, is labeled by its frequency f_k.

8.6.2. Pianos. We now move from calendars to pianos. For more on the interaction between continued fractions and pianos, see [**62**], [**134**], [**15**], [**89**], [**93**], [**9**], [**197**]. Let's start by giving a short lesson on music based on Euler's letter to a German princess [**39**] (see also [**105**]). When, say a piano wire or guitar string vibrates, it causes the air molecules around it to vibrate and these air molecules cause neighboring molecules to vibrate and finally, these molecules bounce against our ears, and we have the sensation of "sound". The rapidness of the vibrations, in number of vibrations per second, is called **frequency**. Let's say that we hear two notes with two different frequencies. In general, these frequencies mix together and don't produce a pleasing sound, but according to Euler, when the *ratio* of their frequencies happens to equal certain ratios of integers, then we hear a pleasant sound![5] Fascinating isn't it? We'll call the ratio of the frequencies an **interval** between the notes or the frequencies. For example, consider two notes, one with frequency f_1 and the other with frequency f_2 such that

$$\frac{f_2}{f_1} = \frac{2}{1} \quad \Longleftrightarrow \quad f_2 = 2f_1 \quad \text{(octave)};$$

in other words, the interval between the first and second note is 2, which is to say, f_2 is just twice f_1. This special interval is called an **octave**. It turns out that when two notes an octave apart are played at the same time, they sound beautiful together! Another interval that is corresponds to a beautiful sound is called the **fifth**, which is when the ratio is 3/2:

$$\frac{f_2}{f_1} = \frac{3}{2} \quad \Longleftrightarrow \quad f_2 = \frac{3}{2}f_1 \quad \text{(fifth)}.$$

Other intervals (which remember just refer to ratios) that have names are

4/3 (fourth) 9/8 (major tone) 25/24 (chromatic semitone),
5/4 (major third) 10/9 (lesser tone) 81/80 (comma of Didymus),
6/5 (minor thirds) 16/15 (diatonic semitone).

However, it is probably of universal agreement that the octave and the fifth make the prettiest sounds. Ratios such as 7/6, 8/7, 11/10, 12/11, ... don't seem to agree with our ears.

Now let's take a quick look at two facts concerning the piano. We all know what a piano keyboard looks like; see Figure 8.1. Let us label the (fundamental) frequencies of the piano keys, counting both white and black, by $f_0, f_1, f_2, f_3, \ldots$

[5] *Musica est exercitium arithmeticae occultum nescientis se numerare animi The pleasure we obtain from music comes from counting, but counting unconsciously. Music is nothing but unconscious arithmetic. From a letter to Goldbach, 27 April 1712, quoted in* [**193**].

starting from the far left key on the keyboard.[6] The first fact is that keys which are twelve keys apart are exactly an octave apart! For instance, f_0 and, jumping twelve keys to the right, f_{12} are an octave apart, f_7 and f_{19} are an octave apart, etc. For this reason, a piano scale really has just twelve basic frequencies, say $f_0, \ldots, f_{11}$, since by doubling these frequencies we get the twelve frequencies above, $f_{12}, \ldots, f_{23}$, and by doubling these we get $f_{24}, \ldots, f_{35}$, etc. The second fact is that a piano is **evenly tempered**, which means that the intervals between adjacent keys is constant. Let this constant be c. Then,

$$\frac{f_{n+1}}{f_n} = c \quad \Longrightarrow \quad f_{n+1} = cf_n$$

for all n. In particular,

$$f_{n+k} = cf_{n+k-1} = c(cf_{n+k-2}) = c^2 f_{n+k-2} = \cdots = c^k f_n. \tag{8.37}$$

Since $f_{n+12} = 2f_n$ (because f_n and f_{n+12} are an octave apart), it follows that with $k = 12$, we get

$$2f_n = c^{12} f_n \quad \Longrightarrow \quad 2 = c^{12} \quad \Longrightarrow \quad c = 2^{1/12}.$$

Thus, the interval between adjacent keys is $2^{1/12}$.

A question that might come to mind is: What is so special about the number twelve for a piano scale? Why not eleven or fifteen? Answer: It has to do with continued fractions! To see why, let us imagine that we have an evenly tempered piano with q basic frequencies, that is, keys that are q apart have frequencies differing by an octave. Question: Which q's make the best pianos? (Note: We better come up with $q = 12$ as one of the "best" ones!) By a very similar argument as we did above, we can see that the interval between adjacent keys is $2^{1/q}$. Now we have to ask: What makes a good piano? Well, our piano by design has octaves, but we would also like our piano to have fifths, the other beautiful interval. Let us label the keys of our piano as in Figure 8.1. Then we would like to have a p such that the interval between any frequency f_n and f_{n+p} is a fifth, that is,

$$\frac{f_{n+p}}{f_n} = \frac{3}{2}.$$

By the formula (8.37), which we can use in the present set-up as long as we put $c = 2^{1/q}$, we have $f_{n+p} = (2^{1/q})^p f_n = 2^{p/q} f_n$. Thus, we want

$$2^{p/q} = \frac{3}{2} \quad \Longrightarrow \quad \frac{p}{q} = \frac{\log(3/2)}{\log 2}.$$

This is, unfortunately, impossible because p/q is rational yet $\frac{\log(3/2)}{\log 2}$ is irrational (cf. Subsection 2.6.5)! Thus, it is impossible for our piano (even if $q = 12$ like our everyday piano) to have a fifth. However, hope is not lost because although our piano can never have a *perfect* fifth, it can certainly have an *approximate* fifth: We just need to find rational approximations to the irrational number $\frac{\log(3/2)}{\log 2}$. This we know how to do using continued fractions. One can show that

$$\frac{\log(3/2)}{\log 2} = \langle 1, 1, 2, 2, 3, 1, \ldots \rangle,$$

[6] A piano wire also gives off **overtones** but we focus here just on the fundamental frequency. Also, some of what we say here is not quite true for the keys near the ends of the keyboard because they don't vibrate well due to their stiffness leading to the phenomenon called **inharmonicity**.

which has convergents

$$0, \frac{1}{1}, \frac{1}{2}, \frac{3}{5}, \frac{7}{12}, \frac{24}{41}, \frac{31}{53}, \frac{179}{306}, \ldots.$$

Lo and behold, we see a twelve! In particular, by the best approximation theorem (Theorem 8.20), we know that 7/12 approximates $\frac{\log(3/2)}{\log 2}$ better than any rational number with a smaller denominator than twelve, which is to say, we cannot find a piano scale with fewer than twelve basic key that will give a better approximation to a fifth. This is why our everyday piano has twelve keys! In summary, $1, 2, 5, 12, 41, 53, 306, \ldots$ are the q's that make the "best" pianos. What about the other numbers in this list? Supposedly [**134**], in 40 B.C. King-Fang, a scholar of the Han dynasty, found the fraction 24/41, although to my knowledge, there has never been an instrument built with a scale of $q = 41$; however, King-Fang also found the fraction 31/53, and in this case, the $q = 53$ scale was advocated by Gerhardus Mercator (1512–1594) circa 1650 and was actually implemented by Robert Halford Macdowall Bosanquet (1841–1912) in his instrument *Enharmonic Harmonium* [**34**]!

We have focused on the interval of a fifth. What about other intervals? ... see Problem 2.

Exercises 8.6.

1. (**Persian calendar**) As of 2000, the modern calendar in Iran and Afghanistan has an average of $365\frac{683}{2820}$ days per year. The persian calendar introduced by Omar Khayyam (1048–1131) had an average of $365\frac{8}{33}$ days per year. Khayyam amazingly calculated the year to be 365.24219858156 days. Find the continued fraction expansion of 365.24219858156 and if $\{c_n\}$ are its convergents, show that c_0 is the ancient calendar, c_1 is the Julian calendar, c_3 is the calendar introduced by Khayyam, and c_7 is the modern Persian calendar!
2. Find the q's that will make a piano with the "best" approximations to a minor third. (Just as we found the q's that will make a piano with the "best" approximations to fifth.) Do you see why many musicians, e.g. Aristoxenus, Kornerup, Ariel, Yasser, who enjoyed minor thirds, liked $q = 19$ musical scales?
3. (**A solar system model**) Christiaan Huygens (1629–1695) made a model scale of the solar system. In his day, it was thought that it took Saturn 29.43 years to make it once around the sun; that is,

$$\frac{\text{period of Saturn}}{\text{period of Earth}} = 29.43.$$

To make a realistic model of the solar system, Huygens needed to make gears for the model Saturn and the model Earth whose number of teeth had a ratio close to 29.43. Find the continued fraction expansion of 29.43 and see why Huygens chose the number of teeth to be 206 and 7, respectively. For more on the use of continued fractions to solve gear problems, see [**147**].

8.7. The elementary functions and the irrationality of $e^{p/q}$

In this section we derive some beautiful and classical continued fraction expansions for $\coth x$, $\tanh x$, and e^x. The book [**126**, Sec. 11.7] has a very nice presentation of this material.

8.7.1. The hypergeometric function.

For complex $a \neq 0, -1, -2, \ldots$, the function

$$F(a,z) := 1 + \frac{1}{a}z + \frac{1}{a(a+1)}\frac{z^2}{2!} + \frac{1}{a(a+1)(a+2)}\frac{z^3}{3!} + \cdots, \quad z \in \mathbb{C},$$

is called a (simplified) **hypergeometric function** or more precisely, the **confluent hypergeometric limit function**. Using the ratio test, it is straightforward to check that $F(a,z)$ converges for all $z \in \mathbb{C}$. If for any $a \in \mathbb{C}$, we define the **pochhammer symbol**, introduced by Leo August Pochhammer (1841–1920),

$$(a)_n := \begin{cases} 1 & n = 0 \\ a(a+1)(a+2)\cdots(a+n-1) & n = 1, 2, 3, \ldots, \end{cases}$$

then we can write the hypergeometric function in shorthand notation:

$$F(a,z) = \sum_{n=0}^{\infty} \frac{1}{(a)_n}\frac{z^n}{n!}.$$

Actually, the true hypergeometric function is defined by (cf. Subsection 6.3.4)

$$F(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n(c)_n}\frac{z^n}{n!},$$

but we won't need this function. Many familiar functions can be written in terms of these hypergeometric functions. For instance, consider

Proposition 8.22. *We have*

$$F\left(\frac{1}{2}, \frac{z^2}{4}\right) = \cosh z \quad , \quad z\,F\left(\frac{3}{2}, \frac{z^2}{4}\right) = \sinh z.$$

Proof. The proof of these identities are the same: We simply check that both sides have the same series expansions. For example, let us check the second identity; the identity for cosh is proved similarly. The function $z\,F\left(\frac{3}{2}, \frac{z^2}{4}\right)$ is just

$$z \cdot \sum_{n=0}^{\infty} \frac{1}{(3/2)_n}\frac{(z^2/2^2)^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{(3/2)_n}\frac{z^{2n+1}}{2^{2n}\,n!},$$

and recall that

$$\sinh z = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}.$$

Thus, we just have to show that $(3/2)_n\,2^{2n}\,n! = (2n+1)!$ for each n. Certainly this holds for $n = 0$. For $n \geq 1$, we have

$$\begin{aligned} (3/2)_n\,2^{2n}\,n! &= \frac{3}{2}\left(\frac{3}{2}+1\right)\left(\frac{3}{2}+2\right)\cdots\left(\frac{3}{2}+n-1\right) \cdot 2^{2n}n! \\ &= \frac{3}{2} \cdot \frac{5}{2} \cdot \frac{7}{2} \cdots \frac{2n+1}{2} \cdot 2^{2n}n! \\ &= 3 \cdot 5 \cdot 7 \cdots (2n+1) \cdot 2^n n! \end{aligned}$$

Since $2^n n! = 2^n \cdot 1 \cdot 2 \cdot 3 \cdots n = 2 \cdot 4 \cdot 6 \cdots 2n$, we have

$$\begin{aligned} 3 \cdot 5 \cdot 7 \cdots (2n+1) \cdot 2^n n! &= 3 \cdot 5 \cdot 7 \cdots (2n+1) \cdot 2 \cdot 4 \cdot 6 \cdots 2n \\ &= 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdots 2n \cdot (2n+1) = (2n+1)! \end{aligned}$$

and our proof is complete. □

The hypergeometric function also satisfies an interesting, and useful as we'll see in a moment, recurrence relation.

PROPOSITION 8.23. *The hypergeometric function satisfies the following recurrence relation:*

$$F(a,z) = F(a+1,z) + \frac{z}{a(a+1)} F(a+2,z).$$

PROOF. The proof of this identity proceeds in the same way as in the previous proposition: We simply check that both sides have the same series expansions. We can write

$$F(a+1,z) + \frac{z}{a(a+1)} F(a+2,z) = \sum_{n=0}^{\infty} \frac{1}{(a+1)_n} \frac{z^n}{n!} + \sum_{n=0}^{\infty} \frac{1}{a(a+1)(a+2)_n} \frac{z^{n+1}}{n!}.$$

The constant term on the right is 1, which is the constant term on the left. For $n \geq 1$, coefficient of z^n on the right is

$$\begin{aligned}
&\frac{1}{(a+1)_n\, n!} + \frac{1}{a(a+1)(a+2)_{n-1}\,(n-1)!} \\
&= \frac{1}{(a+1)\cdots(a+1+n-1)\,n!} + \frac{1}{a(a+1)\cdots(a+2+(n-1)-1)\,(n-1)!} \\
&= \frac{1}{(a+1)\cdots(a+n)\,n!} \;+\; \frac{1}{a(a+1)\cdots(a+n)\,(n-1)!} \\
&= \frac{1}{(a+1)\cdots(a+n)\,(n-1)!} \cdot \left(\frac{1}{n} + \frac{1}{a}\right) \\
&= \frac{1}{(a+1)\cdots(a+n)\,(n-1)!} \left(\frac{a+n}{a \cdot n}\right) \\
&= \frac{1}{a(a+1)\cdots(a+n-1)\,n(n-1)!} = \frac{1}{(a)_n\, n!},
\end{aligned}$$

which is exactly the coefficient of z^n for $F(a,z)$. □

8.7.2. Continued fraction expansion of the hyperbolic cotangent. It turns out that Propositions 8.22 and 8.23 can be combined to give a fairly simple proof of the continued fraction expansion of the hyperbolic cotangent.

THEOREM 8.24. *For any real x, we have*

$$\boxed{\coth x = \frac{1}{x} + \cfrac{x}{3 + \cfrac{x^2}{5 + \cfrac{x^2}{7 + \cfrac{x^2}{9 + \ddots}}}}.}$$

PROOF. With $z = x > 0$, we have $F(a,x) > 0$ for any $a > 0$ by definition of the hypergeometric function. In particular, for $a > 0$, $F(a+1,x) > 0$, so we can divide by this in Proposition 8.23, obtaining the recurrence relation

$$\frac{F(a,x)}{F(a+1,x)} = 1 + \frac{x}{a(a+1)} \frac{F(a+2,x)}{F(a+1,x)},$$

which we can write as

$$\frac{aF(a,x)}{F(a+1,x)} = a + \cfrac{x}{\cfrac{(a+1)F(a+1,x)}{F(a+2,x)}}.$$

Replacing a with $a+n$ with $n = 0,1,2,3,\ldots$, we get

$$\frac{(a+n)F(a+n,x)}{F(a+n+1,x)} = a+n + \cfrac{x}{\cfrac{(a+n+1)F(a+n+1,x)}{F(a+n+2,x)}};$$

that is, if we define

$$\xi_n(a,x) := \frac{(a+n)F(a+n,x)}{F(a+n+1,x)} \quad , \quad a_n := a+n \ , \ b_n := x,$$

then

$$\xi_n(a,x) = a_n + \frac{b_{n+1}}{\xi_{n+1}(a,x)}, \quad n = 0,1,2,3,\ldots. \tag{8.38}$$

Since

$$\sum_{n=1}^{\infty} \frac{a_n a_{n+1}}{b_n} = \sum_{n=1}^{\infty} \frac{(a+n)(a+n+1)}{x} = \infty,$$

by the continued fraction convergence theorem (Theorem 8.14), we know that

$$\frac{aF(a,x)}{F(a+1,x)} = \xi_0(a,x) = a + \frac{x}{a+1+}\,\frac{x}{a+2+}\,\frac{x}{a+3+}\,\frac{x}{a+4+}\,\frac{x}{a+5+}\cdots.$$

Since $F\left(1/2, x^2/4\right) = \cosh x$ and $x\,F\left(3/2, x^2/4\right) = \sinh x$ by Proposition 8.22, when we set $a = 1/2$ and replace x with $x^2/4$ into the previous continued fraction, we find

$$\frac{x\cosh x}{2\sinh x} = \frac{x}{2}\coth x = \frac{1}{2} + \frac{x^2/4}{3/2\ +}\,\frac{x^2/4}{5/2\ +}\,\frac{x^2/4}{7/2\ +}\,\frac{x^2/4}{9/2\ +}\cdots,$$

or after multiplication by 2 and dividing by x, we get

$$\coth x = \frac{1}{x} + \frac{x/2}{3/2+}\,\frac{x^2/4}{5/2\ +}\,\frac{x^2/4}{7/2\ +}\,\frac{x^2/4}{9/2\ +}\cdots,$$

Finally, using the transformation rule (Theorem 8.1)

$$a_0 + \frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\cdots\frac{b_n}{+\,a_n+}\cdots = a_0 + \frac{\rho_1 b_1}{\rho_1 a_1+}\,\frac{\rho_1\rho_2 b_2}{\rho_2 a_2\ +}\cdots+\,\frac{\rho_{n-1}\rho_n b_n}{\rho_n a_n\ \ +}\cdots$$

with $\rho_n = 2$ for all n, we get

$$\coth x = \frac{1}{x} + \frac{x}{3+}\,\frac{x^2}{5\ +}\,\frac{x^2}{7\ +}\,\frac{x^2}{9\ +}\cdots,$$

exactly what we set out to prove. □

Given any x, we certainly have $0 < b_n = x^2 < 2n+1 = a_n$ for all n sufficiently large, so by Theorem 8.15, it follows that when x is rational, $\coth x$ is irrational, or writing it out, for x rational,

$$\coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}} = \frac{e^{2x}+1}{e^{2x}-1}$$

is irrational. It follows that for x rational, e^{2x} must be irrational too, for otherwise $\coth x$ would be rational contrary to assumption. Replacing x with $x/2$ and calling this r, we get the following neat corollary.

THEOREM 8.25. *e^r is irrational for any rational r.*

By the way, as did Johann Heinrich Lambert (1728–1777) originally did back in 1761 [**36**, p. 463], you can use continued fractions to prove that π is irrational, see [**127**], [**154**]. As another easy corollary, we can get the continued fraction expansion for $\tanh x$. To do so, multiply the continued fraction for $\coth x$ by x:

$$x \coth x = b \quad , \quad \text{where } b = 1 + \frac{x^2}{3\,+}\,\frac{x^2}{5\,+}\,\frac{x^2}{7\,+}\,\frac{x^2}{9\,+}\,\cdots.$$

Thus, $\tanh x = \frac{x}{b}$, or replacing b with its continued fraction, we get

$$\boxed{\tanh x = \cfrac{x}{1 + \cfrac{x^2}{3 + \cfrac{x^2}{5 + \cfrac{x^2}{7 + \ddots}}}}\,.}$$

We derive one more beautiful expression that we'll need later. As before, we have

$$\coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}} = \frac{e^{2x}+1}{e^{2x}-1} = \frac{1}{x} + \frac{x}{3+}\,\frac{x^2}{5\,+}\,\frac{x^2}{7\,+}\,\frac{x^2}{9\,+}\,\cdots.$$

Replacing x with $1/x$, we obtain

$$\frac{e^{2/x}+1}{e^{2/x}-1} = x + \frac{1/x}{3\ +}\,\frac{1/x^2}{5\ \ +}\,\frac{1/x^2}{7\ \ +}\,\frac{1/x^2}{9\ \ +}\,\cdots.$$

Finally, using the now familiar transformation rule, after a little algebra we get

$$\boxed{\frac{e^{2/x}+1}{e^{2/x}-1} = x + \cfrac{1}{3x + \cfrac{1}{5x + \cfrac{1}{7x + \ddots}}}\,.} \tag{8.39}$$

8.7.3. Continued fraction expansion of the exponential. We can now get the famous continued fraction expansion for e^x, which was first discovered by (as you might have guessed) Euler. To start, we observe that

$$\coth(x/2) = \frac{e^{x/2} + e^{-x/2}}{e^{x/2} + e^{-x/2}} = \frac{1 + e^{-x}}{1 - e^{-x}} \quad \Longrightarrow \quad e^{-x} = \frac{\coth(x/2) - 1}{1 + \coth(x/2)},$$

where we solved the equation on the left for e^{-x}. Thus,

$$e^{-x} = \frac{\coth(x/2) - 1}{1 + \coth(x/2)} = \frac{1 + \coth(x/2) - 2}{1 + \coth(x/2)} = 1 - \frac{2}{1 + \coth(x/2)},$$

so taking reciprocals, we get

$$e^x = \cfrac{1}{1 - \cfrac{2}{1 + \coth(x/2)}},$$

By Theorem 8.24, we have

$$1+\coth(x/2) = 1 + \frac{2}{x} + \frac{x/2}{3\ +}\,\frac{x^2/4}{5\ +}\cdots = \frac{x+2}{x} + \frac{x/2}{3\ +}\,\frac{x^2/4}{5\ +}\,\frac{x^2/4}{7\ +}\cdots,$$

so

$$e^x = \frac{1}{1+}\,\frac{-2}{\frac{x+2}{x}+}\,\frac{x/2}{3\ +}\,\frac{x^2/4}{5\ +}\,\frac{x^2/4}{7\ +}\cdots$$

or using the transformation rule (Theorem 8.1)

$$\frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\cdots\frac{b_n}{+\,a_n+}\cdots = \frac{\rho_1 b_1}{\rho_1 a_1+}\,\frac{\rho_1\rho_2 b_2}{\rho_2 a_2\ +}\cdots+\,\frac{\rho_{n-1}\rho_n b_n}{\rho_n a_n\quad +}\cdots$$

with $\rho_1 = 1$, $\rho_2 = x$, and $\rho_n = 2$ for all $n \geq 3$, we get

$$e^x = \frac{1}{1+}\,\frac{-2x}{x+2+}\,\frac{x^2}{6\ +}\,\frac{x^2}{10+}\,\frac{x^2}{14+}\cdots.$$

Thus, we have derived Euler's celebrated continued fraction expansion for e^x:

Theorem 8.26. *For any real x, we have*

$$\boxed{e^x = \cfrac{1}{1 - \cfrac{2x}{x + 2 + \cfrac{x^2}{6 + \cfrac{x^2}{10 + \cfrac{x^2}{14 + \ddots}}}}}.}$$

In particular, if we let $x = 1$, we obtain

$$e = \cfrac{1}{1 - \cfrac{2}{3 + \cfrac{1}{6 + \cfrac{1}{10 + \cfrac{1}{14 + \ddots}}}}}.$$

Although beautiful, we can get an even more beautiful continued fraction expansion for e, which is a *simple* continued fraction.

8.7.4. The simple continued fraction expansion of e. If we expand the decimal number 2.718281828 into a simple continued fraction, we get (see Problem 2 in Exercises 8.4)

$$2.718281828 = \langle 2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1\rangle.$$

For this reason, we should be able to conjecture that e is the continued fraction

$$\boxed{e = \langle 2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, \ldots\rangle.} \tag{8.40}$$

This is true, and it was first proved by (as you might have guessed) Euler. Here,

$$a_0 = 2\ ,\ a_1 = 1\ ,\ a_2 = 2\ ,\ a_3 = 1\ ,\ a_4 = 1\ ,\ a_5 = 4\ ,\ a_6 = 1\ ,\ a_7 = 1,$$

and in general, for all $n \in \mathbb{N}$, $a_{3n-1} = 2n$ and $a_{3n} = a_{3n+1} = 1$. Since

$$2 = 1 + \cfrac{1}{0 + \cfrac{1}{1}},$$

we can write (8.40) in a prettier way that shows the full pattern:

$$\boxed{e = \langle 1; 0, 1, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, \ldots \rangle}, \tag{8.41}$$

or in the expanded form

$$\boxed{e = 1 + \cfrac{1}{0 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \ddots}}}}}}}.} \tag{8.42}$$

To prove this incredible formula, denote the convergents of the right-hand continued fraction in (8.40) by r_k/s_k. Since we have such simple relations $a_{3n-1} = 2n$ and $a_{3n} = a_{3n+1} = 1$ for all $n \in \mathbb{N}$, one might think that it is quite easy to compute formulas for r_{3n+1} and s_{3n+1}, and this thought is indeed the case.

LEMMA 8.27. *For all $n \geq 2$, we have*

$$\begin{aligned} r_{3n+1} &= 2(2n+1) r_{3(n-1)+1} + r_{3(n-2)+1} \\ s_{3n+1} &= 2(2n+1) s_{3(n-1)+1} + s_{3(n-2)+1} \end{aligned}$$

PROOF. Both formulas are proved in similar ways, so we shall focus on the formula for r_{3n+1}. First, we apply our Wallis-Euler recursive formulas:

$$r_{3n+1} = r_{3n} + r_{3n-1} = \big(r_{3n-1} + r_{3n-2}\big) + r_{3n-1} = 2r_{3n-1} + r_{3n-2}.$$

We again apply the Wallis-Euler recursive formula on r_{3n-1}:

$$\begin{aligned} r_{3n+1} &= 2\Big(2n r_{3n-2} + r_{3n-3}\Big) + r_{3n-2} \\ &= \Big(2(2n) + 1\Big) r_{3n-2} + 2r_{3n-3} \\ &= \Big(2(2n) + 1\Big) r_{3n-2} + r_{3n-3} + r_{3n-3}. \qquad (8.43) \end{aligned}$$

Again applying the Wallis-Euler recursive formula on the last term, we get

$$\begin{aligned} r_{3n+1} &= \Big(2(2n) + 1\Big) r_{3n-2} + r_{3n-3} + \Big(r_{3n-4} + r_{3n-5}\Big) \\ &= \Big(2(2n) + 1\Big) r_{3n-2} + \Big(r_{3n-3} + r_{3n-4}\Big) + r_{3n-5}. \end{aligned}$$

Since $r_{3n-2} = r_{3n-3} + r_{3n-4}$ by our Wallis-Euler recursive formulas, we finally get

$$\begin{aligned} r_{3n+1} &= \Big(2(2n)+1\Big) r_{3n-2} + r_{3n-2} + r_{3n-5} \\ &= \Big(2(2n)+2\Big) r_{3n-2} + r_{3n-5} \\ &= 2\Big((2n)+1\Big) r_{3(n-1)+1} + r_{3(n-2)+1}. \end{aligned}$$

□

Now putting $x = 1$ in (8.39), let us look at

$$\frac{e+1}{e-1} = \langle 2; 6, 10, 14, 18, \ldots \rangle.$$

that is, if the right-hand side is $\langle \alpha_0; \alpha_1, \ldots \rangle$, then $\alpha_n = 2(2n+1)$ for all $n = 0, 1, 2, \ldots$. If p_n/q_n are the convergents of this continued fraction, then we see that

$$p_n = 2(2n+1)p_{n-1} + p_{n-2} \quad \text{and} \quad q_n = 2(2n+1)q_{n-1} + q_{n-2},$$

which are similar to the relations in our lemma! Thus, it is not surprising in one bit that the r_{3n+1}'s and s_{3n+1}'s are related to the p_n's and q_n's. The exact relation is given in the following lemma.

Lemma 8.28. *For all $n = 0, 1, 2, \ldots$, we have*

$$r_{3n+1} = p_n + q_n \quad \textit{and} \quad s_{3n+1} = p_n - q_n.$$

Proof. As with the previous lemma, we shall only prove the formula for r_{3n+1}. We proceed by induction: First, for $n = 0$, we have

$$r_1 := a_0 a_1 + 1 = 2 \cdot 1 + 1 = 3,$$

while $p_0 := 2$ and $q_0 := 1$, so $r_1 = p_0 + q_0$. If $n = 1$, then by the formula (8.43), which holds for $n \geq 1$, we see that

$$r_{3\cdot 1+1} = (2(2)+1)r_1 + 2r_0 = 5 \cdot 3 + 2 \cdot 2 = 19.$$

On the other hand,

$$p_1 := \alpha_0 \alpha_1 + 1 = 2 \cdot 6 + 1 = 13 \quad , \quad q_1 := \alpha_1 = 6,$$

so $r_{3\cdot 1+1} = p_1 + q_1$.

Assume now that $r_{3k+1} = p_k + q_k$ for all $0 \leq k \leq n-1$ where $n \geq 2$; we shall prove that it holds for $k = n$ (this is an example of "strong induction"; see Section 2.2). But, by Lemma 8.27 and the induction hypothesis, we have

$$\begin{aligned} r_{3n+1} &= 2(2n+1) r_{3(n-1)+1} + r_{3(n-2)+1} \\ &= 2(2n+1)(p_{n-1} + q_{n-2}) + (p_{n-2} + q_{n-2}) \\ &= 2(2n+1)p_{n-1} + p_{n-2} + 2(2n+1)q_{n-2} + q_{n-2} \\ &= p_n + q_n, \end{aligned}$$

where at the last step we used the Wallis-Euler recursive formulas. □

Finally, we can now prove the continued fraction expansion for e:

$$\begin{aligned}\langle 2; 1, 1, 4, 1, 1, \ldots\rangle = \lim \frac{r_n}{s_n} = \lim \frac{r_{3n+1}}{s_{3n+1}} &= \lim \frac{p_n + q_n}{p_n - q_n} \\ &= \lim \frac{p_n/q_n + 1}{p_n/q_n - 1} = \frac{\frac{e+1}{e-1} + 1}{\frac{e+1}{e-1} - 1} = \frac{\frac{e}{e-1}}{\frac{1}{e-1}} = e.\end{aligned}$$

See [**173**] for another proof of this formula based on a proof by Charles Hermite (1822–1901). In the problems, we derive, along with other things, the following beautiful continued fraction for $\cot x$:

$$\boxed{\cot x = \frac{1}{x} + \cfrac{x}{3 - \cfrac{x^2}{5 - \cfrac{x^2}{7 - \cfrac{x^2}{9 - \ddots}}}}\,.} \tag{8.44}$$

From this continued fraction, we can derive the beautiful companion result for $\tan x$:

$$\boxed{\tan x = \cfrac{x}{1 - \cfrac{x^2}{3 - \cfrac{x^2}{5 - \cfrac{x^2}{7 - \ddots}}}}\,.}$$

Exercises 8.7.

1. For all $n = 1, 2, \ldots$, let $a_n > 0$, $b_n \geq 0$, with $a_n \geq b_n + 1$. We shall prove that the following continued fraction converges:

$$\frac{b_1}{a_1 +}\ \frac{-b_2}{a_2\ +}\ \frac{-b_3}{a_3\ +}\ \frac{-b_4}{a_4\ +}\ \cdots. \tag{8.45}$$

Note that for the continued fraction we are studying, $a_0 = 0$. Replacing b_n with $-b_n$ with $n \geq 2$ in the Wallis-Euler recurrence relations (8.16) and (8.17) we get

$$\begin{aligned}&p_n = a_n p_{n-1} - b_n p_{n-2} \quad , \quad q_n = a_n q_{n-1} - b_n q_{n-2}, \quad n = 2, 3, 4, \ldots \\ &p_0 = 0 \quad , \quad p_1 = b_1 \quad , \quad q_0 = 1 \quad , \quad q_1 = a_1.\end{aligned}$$

   (i) Prove (via induction for instance) that $q_n \geq q_{n-1}$ for all $n = 1, 2, \ldots$. In particular, since $q_0 = 1$, we have $q_n \geq 1$ for all n, so the convergents $c_n = p_n/q_n$ of (8.45) are defined.
   (ii) Verify that $q_1 - p_1 \geq 1 = q_0 - p_0$. Now prove by induction that $q_n - p_n \geq q_{n-1} - p_{n-1}$ for all $n = 1, 2, \ldots$. In particular, since $q_0 - p_0 = 1$, we have $q_n - p_n \geq 1$ for all n. Diving by q_n conclude that $0 \leq c_n \leq 1$ for all $n = 1, 2, \ldots$.
   (iii) Using the fundamental recurrence relations for $c_n - c_{n-1}$, prove that $c_n - c_{n-1} \geq 0$ for all $n = 1, 2, \ldots$. Combining this with (ii) shows that $0 \leq c_1 \leq c_2 \leq c_3 \leq \cdots \leq 1$; that is, $\{c_n\}$ is a bounded monotone sequence and hence converges. Thus, the continued fraction (8.45) converges.

2. For all $n = 1, 2, \ldots$, let $a_n > 0$, $b_n \geq 0$, with $a_n \geq b_n + 1$. From the previous problem, it follows that given any $a_0 \in \mathbb{R}$, the continued fraction $a_0 - \frac{b_1}{a_1 +} \frac{-b_2}{a_2\ +} \frac{-b_3}{a_3\ +} \frac{-b_4}{a_4\ +} \cdots$ converges. We now prove a variant of the continued fraction convergence theorem

(Theorem 8.14): Let $\xi_0, \xi_1, \xi_2, \ldots$ be any sequence of real numbers with $\xi_n > 0$ for $n \geq 1$ and suppose that these numbers are related by

$$\xi_n = a_n + \frac{-b_{n+1}}{\xi_{n+1}} \quad , \quad n = 0, 1, 2, \ldots.$$

Then ξ_0 is equal to the continued fraction

$$\xi_0 = a_0 - \frac{b_1}{a_1 +} \, \frac{-b_2}{a_2 \; +} \, \frac{-b_3}{a_3 \; +} \, \frac{-b_4}{a_4 \; +} \, \frac{-b_5}{a_5 \; +} \cdots.$$

Prove this statement following (almost verbatim!) the proof of Theorem 8.14.

3. We are now ready to derive the beautiful cotangent continued fraction (8.44).
   (i) Let $a > 0$. Then as we derived the identity (8.38) found in Theorem 8.24, prove that if we define
   $$\eta_n(a, x) := \frac{(a+n)F(a+n, -x)}{F(a+n+1, -x)} \quad , \quad a_n = a + n \ , \ b_n = x, \qquad n = 0, 1, 2, \ldots,$$
   then
   $$\eta_n(a, x) = a_n + \frac{-b_{n+1}}{\eta_{n+1}(a, x)}, \qquad n = 0, 1, 2, 3, \ldots.$$
   (ii) Using Problem 2, prove that for $x \geq 0$ sufficiently small, we have
   $$\frac{aF(a, -x)}{F(a+1, -x)} = \eta_0(a, x) = a - \frac{x}{a+1+} \, \frac{-x}{a+2+} \, \frac{-x}{a+3+} \, \frac{-x}{a+4+} \, \frac{-x}{a+5+} \cdots. \tag{8.46}$$
   (iii) Prove that (cf. the proof of Proposition 8.22)
   $$F\left(\frac{1}{2}, -\frac{x^2}{4}\right) = \cos x \quad , \quad x\, F\left(\frac{3}{2}, -\frac{x^2}{4}\right) = \sin x.$$
   (iv) Now put $a = 1/2$ and replace x with $-x^2/4$ in (8.46) to derive the beautiful cotangent expansion (8.44). Finally, relax and contemplate this fine formula!
4. (**Irrationality of** $\log r$) Using Theorem 8.25, prove that if $r > 0$ is rational with $r \neq 1$, then $\log r$ is irrational. In particular, one of our favorite constants, $\log 2$, is irrational.

8.8. Quadratic irrationals and periodic continued fractions

We already know (Section 3.8) that a real number has a periodic decimal expansion if and only if the number is rational. One can ask the same thing about continued fractions: What types of real numbers have periodic simple continued fractions? The answer, as you will see in this section, are those real numbers called quadratic irrationals.

8.8.1. Periodic continued fractions. The object of this section is to characterize continued fractions that "repeat".

Example 8.25. We have already encountered the beautiful continued fraction

$$\frac{1+\sqrt{5}}{2} = \langle 1; 1, 1, 1, 1, 1, 1, 1, 1, \ldots \rangle.$$

We usually write the right-hand side as $\langle \overline{1} \rangle$ to emphasize that the 1 repeats.

Example 8.26. Another continued fraction that repeats is

$$\sqrt{8} = \langle 2; 1, 4, 1, 4, 1, 4, 1, 4, \ldots \rangle,$$

where we have an infinite repeating block of $1, 4$. We usually write the right-hand side as $\sqrt{8} = \langle 2; \overline{1, 4} \rangle$.

Example 8.27. Yet one more continued fraction that repeats is

$$\sqrt{19} = \langle 4; 2, 1, 3, 1, 2, 8, 2, 1, 3, 1, 2, 8, \ldots \rangle,$$

where we have an infinite repeating block of $2, 1, 3, 1, 2, 8$. We usually write the right-hand side as $\sqrt{19} = \langle 4; \overline{2, 1, 3, 1, 2, 8} \rangle$.

Notice that the above repeating continued fractions are continued fractions for expressions with square roots.

Example 8.28. Consider now the expression:

$$\xi = \langle 3; 2, 1, 2, 1, 2, 1, 2, 1, \ldots \rangle = \langle 3; \overline{2, 1} \rangle.$$

If $\eta = \langle 2; 1, 2, 1, 2, 1, 2, \ldots \rangle$, then $\xi = 3 + \frac{1}{\eta}$, and

$$\eta = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cdots}}} \quad \Longrightarrow \quad \eta = 2 + \cfrac{1}{1 + \cfrac{1}{\eta}} .$$

Solving for η we get a quadratic formula and solving it, we find that $\eta = 1 + \sqrt{3}$. Hence,

$$\xi = 3 + \frac{1}{\eta} = 3 + \frac{1}{1 + \sqrt{3}} = 3 + \frac{\sqrt{3} - 1}{2} = \frac{5 + \sqrt{3}}{6},$$

yet another square root expression.

Consider the infinite repeating simple continued fraction

$$\begin{aligned} \xi &= \langle a_0; a_1, \ldots, a_{\ell-1}, b_0, b_1, \ldots, b_{m-1}, b_0, b_1, \ldots, b_{m-1}, b_0, b_1, \ldots, b_{m-1}, \ldots \rangle \\ &= \langle a_0; a_1, \ldots, a_{\ell-1}, \overline{b_0, b_1, \ldots, b_{m-1}} \rangle, \end{aligned} \tag{8.47}$$

where the bar denotes that the block of numbers $b_0, b_1, \ldots, b_{m-1}$ repeats forever. Such a continued fraction is said to be **periodic**. When writing a continued fraction in this way we assume that there is no shorter repeating block and that the repeating block cannot start at an earlier position. For example, we would *never* write

$$\langle 2; 1, 2, 4, 3, 4, 3, 4, 3, 4, \ldots \rangle \quad \text{as} \quad \langle 2; 1, 2, 4, \overline{3, 4, 3, 4} \rangle;$$

we simply write it as $\langle 2; 1, 2, \overline{4, 3} \rangle$. The integer m is called the **period** of the simple continued fraction. An equivalent way to define a periodic continued fraction is as an infinite simple continued fraction $\xi = \langle a_0; a_1, a_2, \ldots \rangle$ such that for some m and ℓ, we have

$$a_n = a_{m+n} \quad \text{for all } n = \ell, \ell + 1, \ell + 2, \ldots. \tag{8.48}$$

The examples above suggest that infinite periodic simple continued fractions are intimately related to expressions with square roots; in fact, these expressions are called quadratic irrationals as we shall see in a moment.

8.8.2. Quadratic irrationals. A **quadratic irrational** is, exactly as its name suggests, an irrational real number that is a solution of a quadratic equation with integer coefficients. Using the quadratic equation, we leave you to show that a quadratic irrational ξ can be written in the form

$$\xi = r + s\sqrt{b} \tag{8.49}$$

where r, s are rational numbers and $b > 0$ is an integer that is not a perfect square (for if b were a perfect square, then $\sqrt{b}$ would be an integer so the right-hand side of ξ would be rational, contradicting that ξ is irrational). Conversely, given *any* real number of the form (8.49), one can check that ξ is a root of the equation

$$x^2 - 2r\,x + (r^2 - s^2 b) = 0.$$

Multiplying both sides of this equation by the common denominator of the rational numbers $2r$ and $r^2 - s^2 b$, we can make the polynomial on the left have integer coefficients. Thus, a real number is a quadratic irrational if and only if it is of the form (8.49). As we shall see in Theorem 8.29 below, it would be helpful to write quadratic irrationals in a certain way. Let ξ take the form in (8.49) with $r = m/n$ and $s = p/q$ where we may assume that $n, q > 0$. Then with the help of some mathematical gymnastics, we see that

$$\xi = \frac{m}{n} + \frac{p\sqrt{b}}{q} = \frac{mq + np\sqrt{b}}{nq} = \frac{mq + \sqrt{bn^2p^2}}{nq} = \frac{mnq^2 + \sqrt{bn^4p^2q^2}}{n^2q^2}.$$

Notice that if we set $\alpha = mnq^2$, $\beta = n^2q^2$ and $d = bn^4p^2q^2$, then $d - \alpha^2 = bn^4p^2q^2 - m^2n^2q^4 = (bn^2p^2 - m^2q^2)(n^2q^2)$ is divisible by $\beta = n^2q^2$. Therefore, we can write any quadratic irrational in the form

$$\xi = \frac{\alpha + \sqrt{d}}{\beta}, \quad \alpha, \beta, d \in \mathbb{Z},\ d > 0 \text{ is not a perfect square, and } \beta \big| (d - \alpha^2).$$

Using this expression as the starting point, we prove the following nice theorem that gives formulas for the convergents of the continued fraction expansion of ξ.

THEOREM 8.29. *Let $\xi = \frac{\alpha+\sqrt{d}}{\beta}$ be a quadratic irrational with complete quotients $\{\xi_n\}$ (with $\xi_0 = \xi$) and partial quotients $\{a_n\}$ where $a_n = \lfloor \xi_n \rfloor$. Then,*

$$\xi_n = \frac{\alpha_n + \sqrt{d}}{\beta_n},$$

where α_n and β_n are integers with $\beta_n > 0$ defined by the recursive sequences

$$\alpha_0 = \alpha\ ,\ \beta_0 = \beta\ ,\ \alpha_{n+1} = a_n\beta_n - \alpha_n\ ,\ \beta_{n+1} = \frac{d - \alpha_{n+1}^2}{\beta_n};$$

moreover, $\beta_n | (d - \alpha_n^2)$ for all n.

PROOF. We first show that all the α_n's and β_n's defined above are integers with β_n never zero and $\beta_n | (d - \alpha_n^2)$. This is automatic with $n = 0$. Assume this is true for n. Then $\alpha_{n+1} = a_n\beta_n - \alpha_n$ is an integer. To see that β_{n+1} is also an integer, observe that

$$\beta_{n+1} = \frac{d - \alpha_{n+1}^2}{\beta_n} = \frac{d - (a_n\beta_n - \alpha_n)^2}{\beta_n} = \frac{d - a_n^2\beta_n^2 + 2a_n\beta_n\alpha_n - \alpha_n^2}{\beta_n}$$
$$= \frac{d - \alpha_n^2}{\beta_n} + 2a_n\alpha_n - a_n^2\beta_n.$$

By induction hypothesis, $(d-\alpha_n^2)/\beta_n$ is an integer and so is $2a_n\alpha_n - a_n^2\beta_n$. Thus, β_{n+1} is an integer too. Moreover, $\beta_{n+1} \neq 0$, because if $\beta_{n+1} = 0$, then we must have $d - \alpha_{n+1}^2 = 0$, which shows that d is a perfect square contrary to our condition on d. Finally, since β_n is an integer and

$$\beta_{n+1} = \frac{d-\alpha_{n+1}^2}{\beta_n} \quad \Longrightarrow \quad \beta_n = \frac{d-\alpha_{n+1}^2}{\beta_{n+1}} \quad \Longrightarrow \quad \beta_{n+1} \big| (d-\alpha_{n+1}^2).$$

Lastly, it remains to prove that the ξ_n's are the complete quotients of ξ. To avoid confusion, for each n let's put $\eta_n = (\alpha_n + \sqrt{d})/\beta_n$; we must show that $\eta_n = \xi_n$ for each n. Note that $\eta_0 = \xi = \xi_0$. Now to prove that $\eta_n = \xi_n$ for $n \geq 1$, we simply use the formula for η_n:

$$\eta_n - a_n = \frac{\alpha_n + \sqrt{d}}{\beta_n} - \frac{\alpha_{n+1} + \alpha_n}{\beta_n} = \frac{\sqrt{d} - \alpha_{n+1}}{\beta_n}$$

where in the middle equality we solved $\alpha_{n+1} = a_n\beta_n - \alpha_n$ for a_n. Rationalizing and using the definition of β_{n+1} and ξ_{n+1}, we obtain

$$\eta_n - a_n = \frac{d-\alpha_{n+1}^2}{\beta_n(\sqrt{d}+\alpha_{n+1})} = \frac{\beta_{n+1}}{\sqrt{d}+\alpha_{n+1}} = \frac{1}{\eta_{n+1}} \quad \Longrightarrow \quad \eta_n = a_n + \frac{1}{\eta_{n+1}}.$$

Using this formula plus induction on $n = 0, 1, 2, \ldots$ (recalling that $\eta_0 = \xi_0$) shows that $\eta_n = \xi_n$ for all n. □

8.8.3. Quadratic irrationals and periodic continued fractions. After one preliminary result, we shall prove that an infinite simple continued fraction is a quadratic irrational if and only if it is periodic. Define

$$\mathbb{Z}[\sqrt{d}] := \{a + b\sqrt{d}\,;\; a, b \in \mathbb{Z}\}$$

and

$$\mathbb{Q}[\sqrt{d}] := \{a + b\sqrt{d}\,;\; a, b \in \mathbb{Q}\}.$$

Given $\xi = a + b\sqrt{d}$ in either $\mathbb{Z}[\sqrt{d}]$ or $\mathbb{Q}[\sqrt{d}]$, we define its **conjugate** by

$$\overline{\xi} := a - b\sqrt{d}.$$

Lemma 8.30. *$\mathbb{Z}[\sqrt{d}]$ is a commutative ring and $\mathbb{Q}[\sqrt{d}]$ is a field, and conjugation preserves the algebraic properties; for example, if $\alpha, \beta \in \mathbb{Q}[\sqrt{d}]$, then*

$$\overline{\alpha \pm \beta} = \overline{\alpha} \pm \overline{\beta}, \quad \overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}, \quad \textit{and} \quad \overline{\alpha/\beta} = \overline{\alpha}/\overline{\beta}.$$

Proof. To prove that $\mathbb{Z}[\sqrt{d}]$ is a commutative ring we just need to prove that it has the same algebraic properties as the integers in that $\mathbb{Z}[\sqrt{d}]$ is closed under addition, subtraction, and multiplication — for more on this definition see our discussion in Subsection 2.3.1. For example, to see that $\mathbb{Z}[\sqrt{d}]$ is closed under multiplication, let $\alpha = a + b\sqrt{d}$ and $\beta = a' + b'\sqrt{d}$ be elements of $\mathbb{Z}[\sqrt{d}]$; then,

$$\alpha\beta = (a + b\sqrt{d})(a' + b'\sqrt{d}) = aa' + bb'd + (ab' + a'b)\sqrt{d}, \tag{8.50}$$

which is also in $\mathbb{Z}[\sqrt{d}]$. Similarly, one can show that $\mathbb{Z}[\sqrt{d}]$ satisfies all the other properties of a commutative ring.

To prove that $\mathbb{Q}[\sqrt{d}]$ is a field we need to prove that it has the same algebraic properties as the rational numbers in that $\mathbb{Q}[\sqrt{d}]$ is closed under addition, multiplication, subtraction, and division (by nonzero elements) — for more on this definition see our discussion in Subsection 2.6.1. For example, to see that $\mathbb{Q}[\sqrt{d}]$ is closed under taking reciprocals, observe that if $\alpha = a + b\sqrt{d} \in \mathbb{Q}[\sqrt{d}]$ is not zero, then

$$\frac{1}{\alpha} = \frac{1}{a+b\sqrt{d}} \cdot \frac{a-b\sqrt{d}}{a-b\sqrt{d}} = \frac{a-b\sqrt{d}}{a^2-b^2d} = \frac{a}{a^2-b^2d} - \frac{b}{a^2-b^2d}\sqrt{d}$$

Note that $a^2 - b^2d \neq 0$ since being zero would imply that $\sqrt{d} = a/b$, a rational number, which by assumption is false. Similarly, one can show that $\mathbb{Q}[\sqrt{d}]$ satisfies all the other properties of a field.

Finally, we need to prove that conjugation preserves the algebraic properties. For example, let's prove the equality $\overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}$, leaving the other properties to you. If $\alpha = a + b\sqrt{d}$ and $\beta = a' + b'\sqrt{d}$, then according to (8.50), we have

$$\overline{\alpha\beta} = aa' + bb'd - (ab' + a'b)\sqrt{d}.$$

On the other hand,

$$\overline{\alpha}\overline{\beta} = (a - b\sqrt{d})(a' - b'\sqrt{d}) = aa' + bb'd - (ab' + a'b)\sqrt{d},$$

which equals $\overline{\alpha\beta}$. □

The following theorem was first proved by Joseph-Louis Lagrange (1736–1813).

Theorem 8.31. *An infinite simple continued fraction is a quadratic irrational if and only if it is periodic.*

Proof. We first prove the "if" part then the "only if" part.

Step 1: Let $\xi = \langle a_0; a_1, \ldots, a_{\ell-1}, \overline{b_0, \ldots, b_m}\rangle$ be periodic and define

$$\eta := \langle b_0; b_1, \ldots, b_m, b_0, b_1, \ldots, b_m, b_0, b_1, \ldots, b_m, \ldots\rangle = \langle b_0; b_1, \ldots, b_m, \eta\rangle,$$

so that $\xi = \langle a_0, a_1, \ldots, a_{\ell-1}, \eta\rangle$. Since $\eta = \langle b_0; b_1, \ldots, b_m, \eta\rangle$, by Theorem 8.4, we have

$$\eta = \frac{\eta s_{m-1} + s_{m-2}}{\eta t_{m-1} + t_{m-2}},$$

where s_n/t_n are the convergents for η. Multiplying both sides by $\eta t_{m-1} + t_{m-2}$, we see that

$$\eta^2 t_{m-1} + \eta t_{m-2} = \eta s_{m-1} + s_{m-2} \quad\Longrightarrow\quad a\,\eta^2 + b\,\eta + c = 0,$$

where $a = t_{m-1}$, $b = t_{m-2} - s_{m-1}$, and $c = -s_{m-2}$. Hence, η is a quadratic irrational. Now using that $\xi = \langle a_0, a_1, \ldots, a_{\ell-1}, \eta\rangle$ and Theorem 8.4, we obtain

$$\xi = \frac{\eta p_{m-1} + p_{m-2}}{\eta q_{m-1} + q_{m-2}},$$

where p_n/q_n are the convergents for ξ. Since η is a quadratic irrational, it follows that ξ is a quadratic irrational since $\mathbb{Q}[\sqrt{d}]$ is a field from Theorem 8.30. Thus, we have proved that periodic simple continued fractions are quadratic irrationals.

Step 2: Now let $\xi = \langle a_0; a_1, a_2, \ldots\rangle$ be a quadratic irrational; we shall prove that its continued fraction expansion is periodic. The trick to prove **Step 2** is to first show that the integers α_n and β_n of the complete quotients of ξ found in Theorem 8.29 are bounded. To implement this idea, let ξ_n be the n-th complete quotient of ξ. Then we can write $\xi = \langle a_0; a_1, a_2, \ldots, a_{n-1}, \xi_n\rangle$, so by Theorem 8.4 we have

$$\xi = \frac{\xi_n p_{n-1} + p_{n-2}}{\xi_n q_{n-1} + q_{n-2}}.$$

Solving for ξ_n, after a little algebra, we find that

$$-\xi_n = \frac{q_{n-2}}{q_{n-1}}\left(\frac{\xi - c_{n-2}}{\xi - c_{n-1}}\right).$$

Since conjugation preserves the algebraic operations by our lemma, we see that

$$-\overline{\xi}_n = \frac{q_{n-2}}{q_{n-1}}\left(\frac{\overline{\xi} - c_{n-2}}{\overline{\xi} - c_{n-1}}\right), \tag{8.51}$$

If $\xi = (\alpha + \sqrt{d})/\beta$, then $\overline{\xi} - \xi = 2\sqrt{d}/\beta \neq 0$. Therefore, since $c_k \to \xi$ as $k \to \infty$, it follows that as $n \to \infty$,

$$\left(\frac{\overline{\xi} - c_{n-2}}{\overline{\xi} - c_{n-1}}\right) \to \left(\frac{\overline{\xi} - \xi}{\overline{\xi} - \xi}\right) = 1.$$

In particular, there is an $N \in \mathbb{N}$ such that for $n > N$, $(\overline{\xi} - c_{n-2})/(\overline{\xi} - c_{n-1}) > 0$. Thus, as $q_k > 0$ for $k \geq 0$, according to (8.51), for $n > N$, we have $-\overline{\xi}_n > 0$. Hence, writing ξ_n, which is positive for $n \geq 1$, as $\xi_n = (\alpha_n + \sqrt{d})/\beta_n$ as shown in Theorem 8.29, it follows that for $n > N$,

$$0 = 0 + 0 < \xi_n + (-\overline{\xi}_n) = 2\frac{\sqrt{d}}{\beta_n}.$$

So, for $n > N$, we have $\beta_n > 0$. Now solving the identity $\beta_{n+1} = \frac{d - \alpha_{n+1}^2}{\beta_n}$ in Theorem 8.29 for d we see that

$$\beta_n \beta_{n+1} + \alpha_{n+1}^2 = d.$$

For $n > N$, both β_n and β_{n+1} are positive, which implies that β_n and $|\alpha_n|$ cannot be too large; for instance, for $n > N$, we must have $0 < \beta_n \leq d$ and $0 \leq |\alpha_n| \leq d$. (For if either β_n or $|\alpha_n|$ were greater than d, then $\beta_n\beta_{n+1} + \alpha_{n+1}^2$ would be strictly larger than d, an impossibility since the sum is supposed to equal d.) In particular, if A is the finite set

$$A = \{(j,k) \in \mathbb{Z} \times \mathbb{Z}\,;\, -d \leq j \leq d\ ,\ 1 \leq k \leq d\},$$

then for the infinitely many $n > N$, the pair (α_n, β_n) is in the finite set A. By the pigeonhole principle, there must be distinct $i, j > N$ such that $(\alpha_j, \beta_j) = (\alpha_k, \beta_k)$. Assume that $j > k$ and let $m := j - k$. Then $j = m + k$, so

$$\alpha_k = \alpha_{m+k} \quad \text{and} \quad \beta_k = \beta_{k+m}.$$

Since $a_k = \lfloor \xi_k \rfloor$ and $a_{m+k} = \lfloor \xi_{m+k} \rfloor$, by Theorem 8.29 we have

$$\xi_k = \frac{\alpha_k + \sqrt{d}}{\beta_k} = \frac{\alpha_{m+k} + \sqrt{d}}{\beta_{m+k}} = \xi_{m+k} \quad \Longrightarrow \quad a_k = \lfloor \xi_k \rfloor = \lfloor \xi_{m+k} \rfloor = a_{m+k}.$$

Thus, using our formulas for α_{k+1} and β_{k+1} from Theorem 8.29, we see that

$$\alpha_{k+1} = a_k\beta_k - \alpha_k = a_{m+k}\beta_{m+k} - \alpha_{m+k} = \alpha_{m+k+1},$$

and

$$\beta_{k+1} = \frac{d - \alpha_{k+1}^2}{\beta_k} = \frac{d - \alpha_{m+k+1}^2}{\beta_{m+k}} = \beta_{m+k+1}.$$

Thus,

$$\xi_{k+1} = \frac{\alpha_{k+1} + \sqrt{d}}{\beta_{k+1}} = \frac{\alpha_{m+k+1} + \sqrt{d}}{\beta_{m+k+1}} = \xi_{m+k+1}$$
$$\Longrightarrow \quad a_{k+1} = \lfloor \xi_{k+1} \rfloor = \lfloor \xi_{m+k+1} \rfloor = a_{m+k+1}.$$

Continuing this process by induction shows that $a_n = a_{m+n}$ for all $n = k, k+1, k+2, k+3, \ldots$. Thus, by the definition of periodicity in (8.48), we see that ξ has a periodic simple continued fraction. □

A periodic simple continued fraction is called **purely periodic** if it is of the form $\xi = \langle \overline{a_0; a_1, \ldots, a_{m-1}} \rangle$.

Example 8.29. The simplest example of such a fraction is the golden ratio:

$$\Phi = \frac{1 + \sqrt{5}}{2} = \langle \overline{1} \rangle = \langle 1; 1, 1, 1, 1, 1, \ldots \rangle.$$

Observe that Φ has the following properties:

$$\Phi > 1 \quad \text{and} \quad \overline{\Phi} = \frac{1 - \sqrt{5}}{2} = -0.618\ldots \quad \Longrightarrow \quad \Phi > 1 \quad \text{and} \quad -1 < \overline{\Phi} < 0.$$

In the following theorem, Evariste Galois'[7] (1811–1832) first publication (at the age of 17), we characterize purely periodic expansions as those quadratic irrationals having these same properties. (Don't believe everything to read about the legendary Galois; see [**189**]. See [**220**] for an introduction to Galois' famous theory.)

Theorem 8.32. *A quadratic irrational ξ is purely periodic if and only if*

$$\xi > 1 \quad \textit{and} \quad -1 < \overline{\xi} < 0.$$

Proof. Assume that $\xi = \langle a_0; \ldots, a_{m-1}, a_0, a_1, \ldots, a_{m-1}, \ldots \rangle$ is purely periodic; we shall prove that $\xi > 1$ and $-1 < \overline{\xi} < 0$. Recall that in general, for any simple continued fraction, $\langle b_0; b_1, b_2, \ldots \rangle$ all the b_n's are positive after b_0. Thus, as a_0 appears again (and again, and again, ...) after the first a_0 in ξ, it follows that $a_0 \geq 1$. Hence, $\xi = a_0 + \frac{1}{\xi_1} > 1$. Now applying Theorem 8.4 to $\langle a_0; \ldots, a_{m-1}, \xi \rangle$, we get

$$\xi = \frac{\xi p_{m-1} + p_{m-2}}{\xi q_{m-1} + q_{m-2}},$$

where p_n/q_n are the convergents for ξ. Multiplying both sides by $\xi q_{m-1} + q_{m-2}$, we obtain

$$\xi^2 q_{m-1} + \xi q_{m-2} = \xi p_{m-1} + p_{m-2} \quad \Longrightarrow \quad f(\xi) = 0,$$

where $f(x) = q_{m-1}x^2 + (q_{m-2} - p_{m-1})x - p_{m-2}$ is a quadratic polynomial. In particular, ξ is a root of f. Taking conjugates, we see that

$$q_{m-1}\xi^2 + (q_{m-2} - p_{m-1})\xi - p_{m-2} = 0 \quad \Longrightarrow \quad q_{m-1}\overline{\xi}^2 + (q_{m-2} - p_{m-1})\overline{\xi} - p_{m-2} = 0,$$

[7] *[From the preface to his final manuscript (Evariste died from a pistol duel at the age of 20)] Since the beginning of the century, computational procedures have become so complicated that any progress by those means has become impossible, without the elegance which modern mathematicians have brought to bear on their research, and by means of which the spirit comprehends quickly and in one step a great many computations. It is clear that elegance, so vaunted and so aptly named, can have no other purpose. ... Go to the roots, of these calculations! Group the operations. Classify them according to their complexities rather than their appearances! This, I believe, is the mission of future mathematicians. This is the road on which I am embarking in this work. Evariste Galois (1811–1832).*

therefore $\overline{\xi}$ is the other root of f. Now $\xi > 1$, so by the Wallis-Euler recurrence relations, $p_n > 0$, $p_n < p_{n+1}$, and $q_n < q_{n+1}$ for all n. Hence,

$$f(-1) = (q_{m-1} - q_{m-2}) + (p_{m-1} - p_{m-2}) > 0 \quad \text{and} \quad f(0) = -p_{m-2} < 0.$$

By the intermediate value theorem $f(x) = 0$ for some $-1 < x < 0$. Since $\overline{\xi}$ is the other root of f we have $-1 < \overline{\xi} < 0$.

Assume now that ξ is a quadratic irrational with $\xi > 1$ and $-1 < \overline{\xi} < 0$; we shall prove that ξ is purely periodic. To do so, we first prove that if $\{\xi_n\}$ are the complete quotients of ξ, then $-1 < \overline{\xi}_n < 0$ for all n. Since $\xi_0 = \xi$, this is already true for $n = 0$ by assumption. Assume this holds for n; then,

$$\xi_n = a_n + \frac{1}{\xi_{n+1}} \quad \Longrightarrow \quad \frac{1}{\overline{\xi}_{n+1}} = \overline{\xi}_n - a_n < -a_n \leq -1 \quad \Longrightarrow \quad \frac{1}{\overline{\xi}_{n+1}} < -1.$$

The inequality $\frac{1}{\overline{\xi}_{n+1}} < -1$ shows that $-1 < \overline{\xi}_{n+1} < 0$ and completes the induction. Now we already know that ξ is periodic, so let us assume sake of contradiction that ξ is not purely periodic, that is, $\xi = \langle a_0; a_1, \ldots, a_{\ell-1}, \overline{a_\ell, \ldots, a_{\ell+m-1}}\rangle$ where $\ell \geq 1$. Then $a_{\ell-1} \neq a_{\ell+m-1}$ for otherwise we could start the repeating block at $a_{\ell-1}$, so

$$\xi_{\ell-1} = a_{\ell-1} + \langle\overline{a_\ell, \ldots, a_{\ell+m-1}}\rangle \neq a_{\ell+m-1} + \langle\overline{a_\ell, \ldots, a_{\ell+m-1}}\rangle = \xi_{\ell+m-1} \tag{8.52}$$

Observe that this expression shows that $\xi_{\ell-1} - \xi_{\ell+m-1} = a_{\ell-1} - a_{\ell+m-1}$ is an integer. In particular, taking conjugates, we see that

$$\overline{\xi}_{\ell-1} - \overline{\xi}_{\ell+m-1} = a_{\ell-1} - a_{\ell+m-1} = \xi_{\ell-1} - \xi_{\ell+m-1}.$$

Now we already proved that $-1 < \overline{\xi}_{\ell-1} < 0$ and $-1 < \overline{\xi}_{\ell+m-1} < 0$, which we write as $0 < -\overline{\xi}_{\ell+m-1} < 1$. Thus,

$$0 - 1 < \overline{\xi}_{\ell-1} + (-\overline{\xi}_{\ell+m-1}) < 0 + 1 \quad \Longrightarrow \quad -1 < \xi_{\ell-1} - \xi_{\ell+m-1} < 1,$$

since $\overline{\xi}_{\ell-1} - \overline{\xi}_{\ell+m-1} = \xi_{\ell-1} - \xi_{\ell+m-1}$. However, we noted that $\xi_{\ell-1} - \xi_{\ell+m-1}$ is an integer, and since the only integer strictly between -1 and 1 is 0, it must be that $\xi_{\ell-1} = \xi_{\ell+m-1}$. However, this contradicts (8.52), and our proof is complete. □

8.8.4. Square roots and periodic continued fractions.

Recall that

$$\sqrt{19} = \langle 4; \overline{2, 1, 3, 1, 2, 8}\rangle;$$

if you didn't notice the beautiful symmetry before, observe that we can write this as $\sqrt{19} = \langle a_0; \overline{a_1, a_2, a_3, a_2, a_1, 2a_0}\rangle$ where the repeating block has a symmetric part and an ending part twice a_0. It turns that any square root has this nice symmetry property. To prove this fact, we first prove the following.

LEMMA 8.33. *If $\xi = \langle\overline{a_0; a_1, \ldots, a_{m-1}}\rangle$ is purely periodic, then $-1/\overline{\xi}$ is also purely periodic of the reversed form: $-1/\overline{\xi} = \langle\overline{a_{m-1}; a_{m-2}, \ldots, a_0}\rangle$.*

PROOF. Writing out the complete quotients $\xi, \xi_1, \xi_2, \ldots, \xi_{m-1}$ of

$$\xi = \langle\overline{a_0; a_1, \ldots, a_{m-1}}\rangle = \langle a_0; a_1, \ldots, a_{m-1}, \xi\rangle$$

we obtain

$$\xi = a_0 + \frac{1}{\xi_1}\ ,\ \xi_1 = a_1 + \frac{1}{\xi_2}\ , \ldots,\ \xi_{m-2} = a_{m-2} + \frac{1}{\xi_{m-1}}\ ,\ \xi_{m-1} = a_{m-1} + \frac{1}{\xi}.$$

Taking conjugates of all of these and listing them in reverse order, we find that

$$\frac{-1}{\overline{\xi}} = a_{m-1} - \overline{\xi}_{m-1}\ ,\ \frac{-1}{\overline{\xi}_{m-1}} = a_{m-2} - \overline{\xi}_{m-2}\ , \ldots,\ \frac{-1}{\overline{\xi}_2} = a_1 - \overline{\xi}_1\ ,\ \frac{-1}{\overline{\xi}_1} = a_0 - \overline{\xi}.$$

Let us define $\eta_0 := -1/\overline{\xi}$, $\eta_1 = -1/\overline{\xi}_{m-1}$, $\eta_2 = -1/\overline{\xi}_{m-2}, \ldots, \eta_{m-1} = -1/\overline{\xi}_1$. Then we can write the previous displayed equalities as

$$\eta_0 = a_{m-1} + \frac{1}{\eta_1}\ ,\ \eta_1 = a_{m-2} + \frac{1}{\eta_2}\ , \ldots\ , \eta_{m-2} = a_1 + \frac{1}{\eta_{m-1}}\ ,\ \eta_{m-1} = a_0 + \frac{1}{\eta_0};$$

in other words, η_0 is just the continued fraction:

$$\eta_0 = \langle a_{m-1}; a_{m-2}, \ldots, a_1, a_0, \eta_0\rangle = \langle\overline{a_{m-1}; a_{m-2}, \ldots, a_1, a_0}\rangle.$$

Since $\eta_0 = -1/\overline{\xi}$, our proof is complete. □

Recall that the continued fraction expansion for $\sqrt{d}$ has the complete quotients ξ_n and partial quotients a_n determined by

$$\xi_n = \frac{\alpha_n + \sqrt{d}}{\beta_n}\quad ,\quad a_n = \lfloor \xi_n \rfloor,$$

where the α_n, β_n's are *integers* given in Theorem 8.29. We are now ready to prove Adrien-Marie Legendre's (1752–1833) famous result.

THEOREM 8.34. *The simple continued fraction of $\sqrt{d}$ has the form*

$$\sqrt{d} = \langle a_0; \overline{a_1, a_2, a_3, \ldots, a_3, a_2, a_1, 2a_0}\rangle.$$

Moreover, $\beta_n \neq -1$ for all n, and $\beta_n = +1$ if and only if n is a multiple of the period of $\sqrt{d}$.

PROOF. Starting the continued fraction algorithm for $\sqrt{d}$, we obtain $\sqrt{d} = a_0 + \frac{1}{\xi_1}$, where $\xi_1 > 1$. Since $\frac{1}{\xi_1} = -a_0 + \sqrt{d}$, we have

$$-\frac{1}{\overline{\xi}_1} = -\big(-a_0 - \sqrt{d}\big) = a_0 + \sqrt{d} > 1, \tag{8.53}$$

so we must have $-1 < \overline{\xi}_1 < 0$. Since both $\xi_1 > 1$ and $-1 < \overline{\xi}_1 < 0$, by Galois' Theorem 8.32, we know that ξ_1 is purely periodic: $\xi_1 = \langle\overline{a_1; a_2, \ldots, a_m}\rangle$. Thus,

$$\sqrt{d} = a_0 + \frac{1}{\xi_1} = \langle a_0; \xi_1\rangle = \langle a_0; \overline{a_1, a_2, \ldots, a_m}\rangle.$$

On the other hand, from (8.53) and from Lemma 8.33, we see that

$$\begin{aligned}\langle 2a_0; a_1, a_2, \ldots, a_m, a_1, a_2, \ldots, a_m, \ldots\rangle = a_0 + \sqrt{d} = -\frac{1}{\overline{\xi}_1} = \langle\overline{a_m; \ldots, a_1}\rangle \\ = \langle a_m, a_{m-1}, a_{m-2}, \ldots, a_1, a_m, a_{m-1}, a_{m-2}, \ldots, a_1, \ldots\rangle.\end{aligned}$$

Comparing the left and right-hand sides, we see that $a_m = 2a_0$, $a_{m-1} = a_1$, $a_{m-2} = a_2$, $a_{m-3} = a_3$, and so forth, therefore,

$$\sqrt{d} = \langle a_0; \overline{a_1, a_2, \ldots, a_m}\rangle = \langle a_0; \overline{a_1, a_2, a_3, \ldots, a_3, a_2, a_1, 2a_0}\rangle.$$

We now prove that β_n never equals -1, and $\beta_n = +1$ if and only if n is a multiple of the period m. By the form of the continued fraction expansion of $\sqrt{d}$ we just derived, observe that for any $n > 0$, the n-th complete quotient ξ_n for $\sqrt{d}$ is purely periodic. In particular, by Galois' Theorem 8.32 we know that

$$n > 1 \quad\Longrightarrow\quad \xi_n > 1 \quad\text{and}\quad -1 < \overline{\xi}_n < 0. \tag{8.54}$$

Now for sake of contradiction, assume that $\beta_n = -1$. Since $\beta_0 = +1$ by definition (see Theorem 8.29), we must have $n > 0$. Then the formula $\xi_n = (\alpha_n + \sqrt{d})/\beta_n$ with $\beta_n = -1$ and (8.54) imply that

$$1 < \xi_n = -\alpha_n - \sqrt{d} \quad \Longrightarrow \quad \alpha_n < -1 - \sqrt{d} \quad \Longrightarrow \quad \alpha_n < 0.$$

On the other hand, (8.54) also implies that

$$-1 < \overline{\xi}_n = -\alpha_n + \sqrt{d} < 0 \quad \Longrightarrow \quad \sqrt{d} < \alpha_n \quad \Longrightarrow \quad 0 < \alpha_n.$$

Since $\alpha_n < 0$ and $\alpha_n > 0$ cannot possibly hold, it follows that $\beta_n = -1$ is impossible.

We now prove that $\beta_n = +1$ if and only if n is a multiple of the period m. Assume first that $\beta_n = 1$. Then $\xi_n = \alpha_n + \sqrt{d}$. By (8.54) we see that

$$-1 < \overline{\xi}_n = \alpha_n - \sqrt{d} < 0 \quad \Longrightarrow \quad \sqrt{d} - 1 < \alpha_n < \sqrt{d}.$$

Since α_n is an integer, and the only integer strictly between $\sqrt{d} - 1$ and $\sqrt{d}$ is $a_0 = \lfloor \sqrt{d} \rfloor$, it follows that $\alpha_n = a_0$, so $\xi_n = a_0 + \sqrt{d}$. Now recalling the expansion $\sqrt{d} = \langle a_0; \overline{a_1, a_2, \ldots, a_m} \rangle$ and the fact that $2a_0 = a_m$, it follows that

$$\begin{aligned} a_0 + \sqrt{d} &= \langle 2a_0; a_1, a_2, \ldots, a_{m-1}, a_m, a_1, a_2, \ldots, a_{m-1}, a_m, \ldots \rangle \\ &= \langle \overline{a_m; a_1, a_2, \ldots, a_{m-1}} \rangle; \end{aligned} \tag{8.55}$$

thus $\xi_n = \langle \overline{a_m; a_1, a_2, \ldots, a_{m-1}} \rangle$. On the other hand, ξ_n is by definition the n-th convergent of

$$\sqrt{d} = \langle a_0; a_1, a_2, \ldots, a_m, a_1, a_2, \ldots, a_m, \ldots \rangle,$$

so writing $n = mj + \ell$ where $j = 0, 1, 2, \ldots$ and $1 \leq \ell \leq m$, going out n slots after a_0, we see that

$$\xi_n = \langle \overline{a_\ell; a_{\ell+1}, a_{\ell+2}, \ldots, a_m, a_1, \ldots, a_{\ell-1}} \rangle.$$

Comparing this with $\xi_n = \langle \overline{a_m; a_1, a_2, \ldots, a_{m-1}} \rangle$, we must have $\ell = m$, so $n = mj + m = m(j+1)$ is a multiple of m.

Assume now that n is a multiple of m; say $n = mk$. Then going out $n = mk$ slots to the right of a_0 in the continued fraction expansion of $\sqrt{d}$ we get $\xi_n = \langle \overline{a_m; a_1, a_2, \ldots, a_{m-1}} \rangle$. Thus, $\xi_n = a_0 + \sqrt{d}$ by (8.55). Since $\xi_n = (\alpha_n + \sqrt{d})/\beta_n$ also, it follows that $\beta_n = 1$ and our proof is complete. □

Exercises 8.8.

1. Find the canonical continued fraction expansions for

$$(a)\ \sqrt{29} \quad , \quad (b)\ \frac{1 + \sqrt{13}}{2} \quad , \quad (c)\ \frac{2 + \sqrt{5}}{3}.$$

2. Find the values of the following continued fractions:

$$(a)\ \langle 3; \overline{2, 6} \rangle \quad , \quad (b)\ \langle \overline{1; 2, 3} \rangle \quad , \quad (c)\ \langle 1; 2, \overline{3} \rangle \quad , \quad (d)\ \langle 2; 5, \overline{1, 3, 5} \rangle.$$

3. Let $m, n \in \mathbb{N}$. Find the quadratic irrational numbers represented by

$$(a)\ \langle \overline{n} \rangle = \langle n; n, n, n, \ldots \rangle \quad , \quad (b)\ \langle \overline{n; 1} \rangle \quad , \quad (c)\ \langle \overline{n; n+1} \rangle \quad , \quad (d)\ \langle m; \overline{n} \rangle.$$

8.9. Archimedes' crazy cattle conundrum and diophantine equations

Archimedes of Syracuse (287–212) was known to think in preposterous proportions. In *The Sand Reckoner* [**159**, p. 420], a fun story written by Archimedes, he concluded that if he could fill the universe with grains of sand, there would be approximately 8×10^{63} grains! According to Pappus of Alexandria (290–350), at one time Archimedes said (see [**58**, p. 15]) "Give me a place to stand on, and I will move the earth!" In the following we shall look at a cattle problem proposed by Archimedes, whose solution involves approximately 8×10^{206544} cattle! If you feel mooooooooved to read more on Achimedes' cattle, see [**155**], [**233**], [**19**], [**248**], and [**135**].

8.9.1. Archimedes' crazy cattle conundrum. Here is a poem written by Archimedes to students at Alexandria in a letter to Eratosthenes of Cyrene (276 B.C.–194 B.C.). (The following is adapted from [**98**], as written in [**19**].)

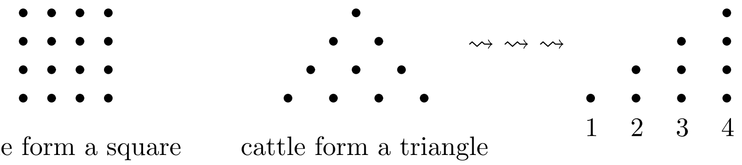
> Compute, O stranger! the number of cattle of Helios, which once grazed on the plains of Sicily, divided according to their color, to wit:
>
> (1) White bulls = $\frac{1}{2}$ black bulls + $\frac{1}{3}$ black bulls + yellow bulls
> (2) Black bulls = $\frac{1}{4}$ spotted bulls + $\frac{1}{5}$ spotted bulls + yellow bulls
> (3) spotted bulls = $\frac{1}{6}$ white bulls + $\frac{1}{7}$ white bulls + yellow bulls
> (4) White cows = $\frac{1}{3}$ black herd + $\frac{1}{4}$ black herd (here, "herd" = bulls + cows)
> (5) Black cows = $\frac{1}{4}$ spotted herd + $\frac{1}{5}$ spotted herd
> (6) Dappled cows = $\frac{1}{5}$ yellow herd + $\frac{1}{6}$ yellow herd
> (7) Yellow cows = $\frac{1}{6}$ white herd + $\frac{1}{7}$ white herd
>
> He who can answer the above is no novice in numbers. Nevertheless he is not yet skilled in wise calculations! But come consider also all the following numerical relations between the Oxen of the Sun:
>
> (8) If the white bulls were combined with the black bulls they would be in a figure equal in depth and breadth and the far stretching plains of Sicily would be covered by the square formed by them.
> (9) Should the yellow and spotted bulls were collected in one place, they would stand, if they ranged themselves one after another, completing the form of an equilateral triangle.
>
> If thou discover the solution of this at the same time; if thou grasp it with thy brain; and give correctly all the numbers; O Stranger! go and exult as conqueror; be assured that thou art by all means proved to have abundant of knowledge in this science.

To solve this puzzle, we need to turn it into mathematics! Let W, X, Y, Z denote the number of white, black, yellow, and spotted bulls, respectively, and w, x, y, z for the number of white, black, yellow, and spotted cows, respectively.

The conditions (1) – (7) can be written as

$$(1)\ W = \Big(\frac{1}{2} + \frac{1}{3}\Big)X + Y \qquad (2)\ X = \Big(\frac{1}{4} + \frac{1}{5}\Big)Z + Y$$

$$(3)\ Z = \Big(\frac{1}{6} + \frac{1}{7}\Big)W + Y \qquad (4)\ w = \Big(\frac{1}{3} + \frac{1}{4}\Big)(X + x)$$



cattle form a square cattle form a triangle

FIGURE 8.2. With the dots as bulls, on the left, the number of bulls is a square number (4^2 in this case) and the number of bulls on the right is a triangular number ($1+2+3+4$ in this case).

$$(5)\ x = \Big(\frac{1}{4}+\frac{1}{5}\Big)(Z+z) \qquad\qquad (6)\ z = \Big(\frac{1}{5}+\frac{1}{6}\Big)(Y+y)$$

$$(7)\ y = \Big(\frac{1}{6}+\frac{1}{7}\Big)(W+w).$$

Now how do we interpret (8) and (9)? We will interpret (8) as meaning that the number of white and black bulls should be a square number (a perfect square); see the left picture in Figure 8.2. A **triangular number** is a number of the form

$$1+2+3+4+\cdots+n = \frac{n(n+1)}{2},$$

for some n. Then we will interpret (9) as meaning that the number of yellow and spotted bulls should be a triangular number; see the right picture in Figure 8.2. Thus, (8) and (9) become

$$(8)\ W+X = \text{a square number} \quad , \qquad (9)\ Y+Z = \text{a triangular number.}$$

In summary: We want to find *integers* W, X, Y, Z, w, x, y, z (here we assume there are no such thing as "fractional cattle") solving equations (1)–(9). Now to the solution of Archimedes cattle problem. First of all, equations (1)–(7) are just linear equations so these equations can be solved using simple linear algebra. Instead of solving these equations by hand, which will probably take a few hours, it might be best to use a computer. Doing so you will find that in order for W, X, Y, Z, w, x, y, z to solve (1)–(7), they must be of the form

$$\begin{array}{ll} (8.56) & \begin{array}{l} W = 10366482\,k \ \ , \ \ X = 7460514\,k \ \ , \ \ Y = 4149387\,k \ \ , \ \ Z = 7358060\,k \\ w = 7206360\,k \ \ , \ \ x = 4893246\,k \ \ , \ \ y = 5439213\,k \ \ , \ \ z = 3515820\,k, \end{array} \end{array}$$

where k can equal $1, 2, 3, \ldots$. Thus, in order for us to fulfill conditions (1)–(7), we would have at the very least, setting $k = 1$,

$$\begin{aligned} &10366482 + 7460514 + 4149387 + 7358060 + 7206360 + 4893246 \\ &\qquad\qquad + 5439213 + 3515820 = 50389082 \approx 50 \text{ million cattle!} \end{aligned}$$

Now we are "no novice in numbers!" Nevertheless we are not yet skilled in wise calculations! To be skilled, we still have to satisfy conditions (8) and (9). For (8), this means

$$W + X = 10366482\,k + 7460514\,k = 17826996\,k = \text{a square number.}$$

Factoring $17826996 = 2^2 \cdot 3 \cdot 11 \cdot 29 \cdot 4657$ into its prime factors, we see that we must have

$$2^2 \cdot 3 \cdot 11 \cdot 29 \cdot 4657\,k = (\cdots)^2,$$

a square of an integer. Thus, we need $3 \cdot 11 \cdot 29 \cdot 4657\, k$ to be a square, which holds if and only if

$$k = 3 \cdot 11 \cdot 29 \cdot 4657\, m^2 = 4456749\, m^2$$

for some integer m. Plugging this value into (8.56), we get

$$\begin{aligned} W &= 46200808287018\, m^2 \quad , \quad X = 33249638308986\, m^2 \\ Y &= 18492776362863\, m^2 \quad , \quad Z = 32793026546940\, m^2 \\ w &= 32116937723640\, m^2 \quad , \quad x = 21807969217254\, m^2 \\ y &= 24241207098537\, m^2 \quad , \quad z = 15669127269180\, m^2, \end{aligned} \tag{8.57}$$

where m can equal $1, 2, 3, \ldots$. Thus, in order for us to fulfill conditions (1)–(8), we would have at the very least, setting $m = 1$,

$$\begin{aligned} &46200808287018 + 33249638308986 + 18492776362863 + 32793026546940 \\ &\quad + 32116937723640 + 21807969217254 + 24241207098537 \\ &\quad\quad + 15669127269180 = 2.2457\ldots \times 10^{14} \approx 2.2 \text{ trillion cattle!} \end{aligned}$$

It now remains to satisfy condition (9):

$$\begin{aligned} Y + Z &= 18492776362863\, m^2 + 32793026546940\, m^2 \\ &= 51285802909803\, m^2 = \frac{\ell(\ell+1)}{2}, \end{aligned}$$

for some integer ℓ. Multiplying both sides by 8 and adding 1, we obtain

$$8 \cdot 51285802909803\, m^2 + 1 = 4\ell^2 + 4\ell + 1 = (2\ell + 1)^2 = n^2,$$

where $n = 2\ell + 1$. Since $8 \cdot 51285802909803 = 410286423278424$, we finally conclude that conditions (1)–(9) are all fulfilled if we can find *integers* m, n satisfying the equation

$$n^2 - 410286423278424\, m^2 = 1. \tag{8.58}$$

This is commonly called a **Pell equation** and is an example of a diophantine equation. As we'll see in the next subsection, we can solve this equation by simply (!) finding the simple continued fraction expansion of $\sqrt{410286423278424}$. The calculations involved are just sheer madness, but they can be done and have been done [**19**], [**248**]. In the end, we find that the smallest total number of cattle which satisfy (1)–(9) is a number with 206545 digits (!) and is equal to

$$7760271406\ldots(206525 \text{ other digits go here})\ldots 9455081800 \approx 8 \times 10^{206544}.$$

We are now skilled in wise calculations! A copy of this number is printed on 42 computer sheets and has been deposited in the Mathematical Tables of the journal *Mathematics of Computation* if you are interested.

8.9.2. Pell's equation. Generalizing the cattle equation (8.58), we call a diophantine equation of the form

$$x^2 - d\, y^2 = 1 \tag{8.59}$$

a **Pell equation**. Note that $(x, y) = (1, 0)$ solves this equation. This solution is called the **trivial solution**; the other solutions are not so easily attained. We remark that Pell's equation was named by Euler after John Pell (1611–1685), although Brahmagupta[8] (598–670) studied this equation a thousand years earlier [**36**, p. 221]. Any case, we shall see that the continued fraction expansion of $\sqrt{d}$ plays an important role in solving this equation. We note that if (x, y) solves (8.59), then trivially so do $(\pm x, \pm y)$ because of the squares in (8.59); thus, we usually restrict ourselves to the positive solutions.

Recall that the continued fraction expansion for $\sqrt{d}$ has the complete quotients ξ_n and partial quotients a_n determined by

$$\xi_n = \frac{\alpha_n + \sqrt{d}}{\beta_n} \quad , \quad a_n = \lfloor \xi_n \rfloor,$$

where α_n and β_n are integers defined in Theorem 8.29. The exact forms of these integers are not important; what is important is that β_n never equals -1 and $\beta_n = +1$ if and only if n is a multiple of the period of $\sqrt{d}$ as we saw in Theorem 8.34. The following lemma shows how the convergents of $\sqrt{d}$ enter Pell's equation.

Lemma 8.35. *If p_n/q_n denotes the n-th convergent of $\sqrt{d}$, then for all $n = 0, 1, 2, \ldots$, we have*

$$p_n^2 - d\,q_n^2 = (-1)^{n+1}\beta_{n+1}.$$

Proof. Since we can write $\sqrt{d} = \langle a_0; a_1, a_2, a_3, \ldots, a_n, \xi_{n+1}\rangle$ and $\xi_{n+1} = (\alpha_{n+1} + \sqrt{d})/\beta_{n+1}$, by (8.19) of Corollary 8.6, we have

$$\sqrt{d} = \frac{\xi_{n+1}p_n + p_{n-1}}{\xi_{n+1}q_n + q_{n-1}} = \frac{(\alpha_{n+1} + \sqrt{d})\,p_n + \beta_{n+1}p_{n-1}}{(\alpha_{n+1} + \sqrt{d})\,q_n + \beta_{n+1}q_{n-1}}.$$

Multiplying both sides by the denominator of the right-hand side, we get

$$\begin{aligned} &\sqrt{d}(\alpha_{n+1} + \sqrt{d})\,q_n + \sqrt{d}\beta_{n+1}q_{n-1} = (\alpha_{n+1} + \sqrt{d})\,p_n + \beta_{n+1}p_{n-1} \\ &\qquad \Longrightarrow \quad dq_n + (\alpha_{n+1}q_n + \beta_{n+1}q_{n-1})\sqrt{d} = (\alpha_{n+1}p_n + \beta_{n+1}p_{n-1}) + p_n\sqrt{d}. \end{aligned}$$

Equating coefficients, we obtain

$$dq_n = \alpha_{n+1}p_n + \beta_{n+1}p_{n-1} \quad \text{and} \quad \alpha_{n+1}q_n + \beta_{n+1}q_{n-1} = p_n.$$

Multiplying the first equation by q_n and the second equation by p_n and equating the $\alpha_{n+1}p_nq_n$ terms in each resulting equation, we obtain

$$\begin{aligned} &dq_n^2 - \beta_{n+1}p_{n-1}q_n = p_n^2 - \beta_{n+1}p_nq_{n-1} \\ &\qquad\qquad \Longrightarrow \quad p_n^2 - d\,q_n^2 = (p_nq_{n-1} - p_{n-1}q_n)\cdot\beta_{n+1} = (-1)^{n+1}\cdot\beta_{n+1}, \end{aligned}$$

where we used that $p_nq_{n-1} - p_{n-1}q_n = (-1)^{n-1} = (-1)^{n+1}$ from Corollary 8.6. □

Next, we show that *all* solutions of Pell's equation can be found via the convergents of $\sqrt{d}$.

Theorem 8.36. *Let p_n/q_n denote the n-th convergent of $\sqrt{d}$ and let m the period of $\sqrt{d}$. Then the positive integer solutions to*

$$x^2 - d\,y^2 = 1$$

[8] *A person who can, within a year, solve $x^2 - 92y^2 = 1$ is a mathematician. Brahmagupta (598–670).*

are precisely numerators and denominators of the odd convergents of $\sqrt{d}$ of the form $x = p_{nm-1}$ and $y = q_{nm-1}$, where $n > 0$ is any positive integer for m even and $n > 0$ is even for m odd.

Proof. We prove our theorem in two steps.

Step 1: We first prove that if $x^2 - d\,y^2 = 1$ with $y > 0$, then x/y is a convergent of $\sqrt{d}$. To see this, observe that since $1 = x^2 - d\,y^2 = (x - \sqrt{d}\,y)(x + \sqrt{d}\,y)$, we have $x - \sqrt{d}\,y = 1/(x + \sqrt{d}\,y)$, so

$$\left|\frac{x}{y} - \sqrt{d}\right| = \left|\frac{x - \sqrt{d}\,y}{y}\right| = \frac{1}{y\,|x + \sqrt{d}\,y|}.$$

Also, $x^2 = d\,y^2 + 1 > d\,y^2$ implies $x > \sqrt{d}\,y$, which implies

$$x + \sqrt{d}\,y > \sqrt{d}\,y + \sqrt{d}\,y = 2\sqrt{d}\,y.$$

Hence,

$$\left|\frac{x}{y} - \sqrt{d}\right| = \frac{1}{y\,|x + \sqrt{d}\,y|} < \frac{1}{y \cdot 2\sqrt{d}\,y} = \frac{1}{2y^2\sqrt{d}} \quad\Longrightarrow\quad \left|\frac{x}{y} - \sqrt{d}\right| < \frac{1}{2y^2}.$$

By Dirichlet's theorem 8.21, it follows that x/y must be a convergent of $\sqrt{d}$.

Step 2: We now finish the proof. By **Step 1** we already know that every solution must be a convergent, so we only need to look for convergents (p_k, q_k) that make $p_k^2 - d\,q_k^2 = 1$. To this end, recall from Lemma 8.35 that

$$p_{k-1}^2 - d\,q_{k-1}^2 = (-1)^k \beta_k,$$

where β_k never equals -1 and $\beta_k = 1$ if and only if k is a multiple of m, the period of $\sqrt{d}$. In particular, $p_{k-1}^2 - d\,q_{k-1}^2 = 1$ if and only if $(-1)^k\beta_k = 1$, if and only if $\beta_k = 1$ and k is even, if and only if k is a multiple of m and k is even. This holds if and only if $k = mn$ where $n > 0$ is any positive integer for m even and $n > 0$ is even for m odd. This completes our proof. □

The **fundamental solution** of Pell's equation is the "smallest" positive solution of Pell's equation; here, a solution (x, y) is **positive** means $x, y > 0$. Explicitly, the fundamental solution is (p_{m-1}, q_{m-1}) for an even period m of $\sqrt{d}$ or (p_{2m-1}, p_{2m-1}) for an odd period m.

Example 8.30. Consider the equation $x^2 - 3y^2 = 1$. Since $\sqrt{3} = \langle 1; \overline{1, 2}\rangle$ has period $m = 2$, our theorem says that the positive solutions of $x^2 - 3y^2 = 1$ are precisely $x = p_{2n-1}$ and $y = q_{2n-1}$ for all $n > 0$; that is, $(p_1, q_1), (p_3, q_3), (p_5, q_5), \ldots$. Now the convergents of $\sqrt{3}$ are

n	0	1	2	3	4	5	6	7
$\frac{p_n}{q_n}$	$\frac{1}{1}$	$\frac{2}{1}$	$\frac{5}{3}$	$\frac{7}{4}$	$\frac{19}{11}$	$\frac{26}{15}$	$\frac{71}{41}$	$\frac{97}{56}$

.

In particular, the fundamental solution is $(2, 1)$ and the rest of the positive solutions are $(7, 4)$, $(26, 15)$, $(97, 56)$, ... Just to verify a couple entries:

$$2^2 - 3 \cdot 1^2 = 4 - 3 = 1$$

and

$$7^2 - 3 \cdot 4^2 = 49 - 3 \cdot 16 = 49 - 48 = 1,$$

and one can continue verifying that the odd convergents give solutions.

Example 8.31. For another example, consider the equation $x^2 - 13\,y^2 = 1$. In this case, we find that $\sqrt{13} = \langle 3; \overline{1,1,1,1,6}\rangle$ has period $m = 5$. Thus, our theorem says that the positive solutions of $x^2 - 13y^2 = 1$ are precisely $x = p_{5n-1}$ and $y = q_{5n-1}$ for all $n > 0$ *even*; that is, $(p_9, q_9), (p_{19}, q_{19}), (p_{29}, q_{29}), \ldots$. The convergents of $\sqrt{13}$ are

n	0	1	2	3	4	5	6	7	8	9
$\frac{p_n}{q_n}$	$\frac{3}{1}$	$\frac{4}{1}$	$\frac{7}{2}$	$\frac{11}{3}$	$\frac{18}{5}$	$\frac{119}{33}$	$\frac{137}{38}$	$\frac{256}{71}$	$\frac{393}{109}$	$\frac{649}{180}$

.

In particular, the fundamental solution is $(649, 180)$.

8.9.3. Brahmagupta's algorithm. Thus, to find solutions of Pell's equation we just have to find certain convergents of $\sqrt{d}$. Finding all convergents is quite a daunting task — try finding the solution (p_{19}, q_{19}) for $\sqrt{13}$ — but it turns out that all the positive solutions can be found from the fundamental solution.

Example 8.32. We know that the fundamental solution of $x^2 - 3y^2 = 1$ is $(2, 1)$ and the rest of the positive solutions are $(7, 4)$, $(26, 15)$, $(97, 56)$, ... Observe that

$$(2 + 1 \cdot \sqrt{3})^2 = 4 + 4\sqrt{3} + 3 = 7 + 4\sqrt{3}.$$

Note that the second positive solution $(7, 4)$ to $x^2 - 3y^2 = 1$ appears on the right. Now observe that

$$(2 + 1 \cdot \sqrt{3})^3 = (2 + \sqrt{3})^2\,(2 + \sqrt{3}) = (7 + 4\sqrt{3})\,(2 + \sqrt{3}) = 26 + 15\sqrt{3}.$$

Note that the third positive solution $(26, 15)$ to $x^2 - 3y^2 = 1$ appears on the right. One may conjecture that the n-th positive solution (x_n, y_n) to $x^2 - 3\,y^2 = 1$ is found by multiplying out

$$x_n + y_n\,\sqrt{d} = (2 + 1 \cdot \sqrt{3})^n$$

This is in fact correct as the following theorem shows.

THEOREM 8.37 (**Brahmagupta's algorithm**). *If (x_1, y_1) is the fundamental solution of Pell's equation*

$$x^2 - d\,y^2 = 1,$$

then all the other positive solutions (x_n, y_n) can be obtained from the equation

$$x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n \quad , \quad n = 0, 1, 2, 3, \ldots.$$

PROOF. To simplify this proof a little, we shall say that $\zeta = x + y\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ solves Pell's equation to mean that (x, y) solves Pell's equation; similarly, we say ζ is a positive solution to mean that $x, y > 0$. Throughout this proof we shall use the following fact:

$$\zeta \text{ solves Pell's equation} \quad \Longleftrightarrow \quad \zeta\,\overline{\zeta} = 1 \;\; (\text{that is, } 1/\zeta = \overline{\zeta}). \tag{8.60}$$

This is holds for the simple reason that

$$\zeta\,\overline{\zeta} = (x + y\sqrt{d})\,(x - y\sqrt{d}) = x^2 - d\,y^2.$$

In particular, if we set $\alpha := x_1 + y_1\sqrt{d}$, then $\alpha\,\overline{\alpha} = 1$ because (x_1, y_1) solves Pell's equation. We now prove our theorem. We first note that (x_n, y_n) is a solution because

$$(x_n + y_n\sqrt{d})\,\overline{(x_n + y_n\sqrt{d})} = \alpha^n \cdot \overline{\alpha^n} = \alpha^n \cdot (\overline{\alpha})^n = (\alpha \cdot \overline{\alpha})^n = 1^n = 1,$$

which in view of (8.60), we conclude that (x_n, y_n) solves Pell's equation. Now suppose that $\xi \in \mathbb{Z}[\sqrt{d}]$ is a positive solution to Pell's equation; we must show that ξ is some power of α. To this end, note that $\alpha \le \xi$ because $\alpha = x_1 + y_1\sqrt{d}$ and (x_1, y_1) is the smallest positive solution of Pell's equation. Since $1 < \alpha$, it follows that $\alpha^k \to \infty$ as $k \to \infty$, so we can choose $n \in \mathbb{N}$ to be the smallest natural number such that $\xi < \alpha^{n+1}$. Then, $\alpha^n \le \xi < \alpha^{n+1}$, so dividing by α^n, we obtain

$$1 \le \eta < \alpha \quad \text{where} \quad \eta := \frac{\xi}{\alpha^n} = \xi \cdot (\overline{\alpha})^n,$$

where we used that $1/\alpha = \overline{\alpha}$ from (8.60). Since $\mathbb{Z}[\sqrt{d}]$ is a ring (Lemma 8.30), we know that $\eta = \xi \cdot (\overline{\alpha})^n \in \mathbb{Z}[\sqrt{d}]$ as well. Moreover, η solves Pell's equation because

$$\eta\,\overline{\eta} = \xi \cdot (\overline{\alpha})^n \cdot \overline{\xi} \cdot \alpha^n = (\xi\,\overline{\xi}) \cdot (\overline{\alpha}\,\alpha)^n = 1 \cdot 1 = 1.$$

We shall prove that $\eta = 1$, which shows that $\xi = \alpha^n$. To prove this, observe that from $1 \le \eta < \alpha$ and the fact that $1/\eta = \overline{\eta}$ (since $\eta\,\overline{\eta} = 1$), we have

$$0 < \alpha^{-1} < \eta^{-1} \le 1 \quad \Longrightarrow \quad 0 < \alpha^{-1} < \overline{\eta} \le 1.$$

Let $\eta = p + q\sqrt{d}$ where $p, q \in \mathbb{Z}$. Then the inequalities $1 \le \eta < \alpha$ and $0 < \alpha^{-1} < \overline{\eta} \le 1$ imply that

$$2p = (p + q\sqrt{d}) + (p - q\sqrt{d}) = \eta + \overline{\eta} \ge 1 + \alpha^{-1} > 0$$

and

$$2q\sqrt{d} = (p + q\sqrt{d}) - (p - q\sqrt{d}) = \eta - \overline{\eta} \ge 1 - 1 = 0.$$

In particular, $p > 0$, $q \ge 0$, and $p^2 - dq^2 = 1$ (since η solves Pell's equation). Therefore, $(p, q) = (1, 0)$ or (p, q) is a positive (numerator, denominator) of a convergent of $\sqrt{d}$. However, we know that (x_1, y_1) is the smallest such positive (numerator, denominator), and that $p + q\sqrt{d} = \eta < \alpha = x_1 + y_1\sqrt{d}$. Therefore, we must have $(p, q) = (1, 0)$. This implies that $\eta = 1$ and hence $\xi = \alpha^n$. □

Example 8.33. Since (649, 180) is the fundamental solution to $x^2 - 13\,y^2 = 1$, all the positive solutions are given by

$$x_n + y_n\sqrt{13} = (649 + 180\,\sqrt{13})^n.$$

For instance, for $n = 2$, we find that

$$(649 + 180\,\sqrt{13})^2 = 842401 + 233640\sqrt{13} \quad \Longrightarrow \quad (x_2, y_2) = (842401, 233640),$$

much easier than finding (p_{19}, q_{19}).

There are many cool applications of Pell's equation explored in the exercises. Here's one of my favorites (see Problem 8): Any prime of the form $p = 4k + 1$ is a sum of two squares. This was conjectured by Pierre de Fermat[9] (1601–1665) in 1640 and proved by Euler in 1754. For example, 5, 13, 17 are such primes, and $5 = 1^2 + 2^2$, $13 = 2^2 + 3^2$, and $17 = 1^2 + 4^2$.

Exercises 8.9.

[9] *[In the margin of his copy of Diophantus' Arithmetica, Fermat wrote] To divide a cube into two other cubes, a fourth power or in general any power whatever into two powers of the same denomination above the second is impossible, and I have assuredly found an admirable proof of this, but the margin is too narrow to contain it. Pierre de Fermat (1601–1665). Fermat's claim in this marginal note, later to be called "Fermat's last theorem" remained an unsolved problem in mathematics until 1995 when Andrew Wiles (1953 –) finally proved it.*

1. Find the fundamental solutions to the equations

$$(a)\ x^2 - 8\,y^2 = 1 \quad , \quad (b)\ x^2 - 5\,y^2 = 1 \quad , \quad (c) x^2 - 7\,y^2 = 1.$$

   Using the fundamental solution, find the next two solutions.

2. (**Pythagorean triples**) (Cf. [**174**]) Here's is a nice problem solvable using continued fractions. A **Pythagorean triple** consists of three natural numbers (x, y, z) such that $x^2 + y^2 = z^2$. For example, $(3, 4, 5)$, $(5, 12, 13)$, and $(8, 15, 17)$ are examples. (Can you find more?) The first example $(3, 4, 5)$ has the property that the first two are consecutive integers; here are some steps to find more Pythagorean triples of this sort.
   (i) Show that (x, y, z) is a Pythagorean triple with $y = x + 1$ if and only if

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

   (ii) By solving the Pell equation $u^2 - 2\,v^2 = 1$, find the next three Pythagorean triples (x, y, z) (after $(3, 4, 5)$) where x and y are consecutive integers.

3. (**Triangular numbers**) Here's is another very nice problem that can be solved using continued fractions. Find all triangular numbers that are squares, where recall that a triangular number is of the form $1 + 2 + \cdots + n = n(n + 1)/2$. Here are some steps.
   (i) Show that $n(n + 1)/2 = m^2$ if and only if

$$(2n + 1)^2 - 8m^2 = 1.$$

   (ii) By solving the Pell equation $x^2 - 8\,y^2 = 1$, find the first three triangular numbers that are squares.

4. In this problem we answer the question: For which $n \in \mathbb{N}$ is the standard deviation of the $2n + 1$ numbers $0, \pm 1, \ldots, \pm n$ an integer? Here, the **standard deviation** of any real numbers $x_1, \ldots, x_N$ is by definition the number $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$ where $\overline{x}$ is the average of $x_1, \ldots, x_N$.
   (i) Show that the standard deviation of $0, \pm 1, \ldots, \pm n$ equals $\sqrt{\frac{1}{3}n(n + 1)}$. Suggestion: The formula $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$ from Problem 3b of Exercises 2.2 might be helpful.
   (ii) Therefore, we want $\frac{1}{3}n(n + 1) = y^2$ where $y \in \mathbb{N}$. If we put $x = 2n + 1$, prove that $\frac{1}{3}n(n + 1) = y^2$ if and only if $x^2 - 12y^2 = 1$ where $x = 2n + 1$.
   (iii) Now solve the equation $x^2 - 12y^2 = 1$ to answer our question.

5. The diophantine equation $x^2 - d\,y^2 = -1$ (where $d > 0$ is not a perfect square) is also of interest. In this problem we determine when this equation has solutions. Following the proof of Theorem 8.36, prove the following statements.
   (i) Show that if (x, y) solves $x^2 - d\,y^2 = -1$ with $y > 0$, then x/y is a convergent of $\sqrt{d}$.
   (ii) Prove that $x^2 - d\,y^2 = -1$ has a solution if and only if the period of $\sqrt{d}$ is odd, in which case the nonnegative solutions are exactly $x = p_{nm-1}$ and $y = q_{nm-1}$ for all $n > 0$ odd.

6. Which of the following equations have solutions? If an equation has solutions, find the fundamental solution.

$$(a)\ x^2 - 2\,y^2 = -1 \quad , \quad (b)\ x^2 - 3\,y^2 = -1 \quad , \quad (c) x^2 - 17\,y^2 = -1.$$

7. In this problem we prove that the diophantine equation $x^2 - p\,y^2 = -1$ always has a solution if p is a prime number of the form $p = 4k + 1$ for an integer k. For instance, since $13 = 4 \cdot 3 + 1$ and $17 = 4 \cdot 4 + 1$, $x^2 - 13y^2 = -1$ and $x^2 - 17y^2 = -1$ have solutions (as you already saw in the previous problem). Let $p = 4k + 1$ be prime.
   (i) Let (x_1, y_1) be the fundamental solution of $x^2 - p\,y^2 = 1$. Prove that x_1 and y_1 cannot both be even and cannot both be odd.
   (ii) Show that the case x_1 is even and y_1 is odd cannot happen. Suggestion: Write $x_1 = 2a$ and $y_1 = 2b + 1$ and plug this into $x_1^2 - p\,y_1^2 = 1$.

(iii) Thus, we may write $x_1 = 2a+1$ and $y_1 = 2b$. Show that $p\,b^2 = a\,(a+1)$. Conclude that p must divide a or $a+1$.

(iv) Suppose that p divides a; that is, $a = mp$ for an integer m. Show that $b^2 = m\,(mp+1)$ and that m and $mp+1$ are relatively prime. Using this equality, prove that $m = s^2$ and $mp+1 = t^2$ for integers s, t. Conclude that $t^2 - p\,s^2 = 1$ and derive a contradiction.

(v) Thus, it must be the case that p divides $a+1$. Using this fact and an argument similar to the one in the previous step, find a solution to $x^2 - d\,y^2 = -1$.

8. (**Sum of squares**) In this problem we prove the following incredible result of Euler: Every prime of the form $p = 4k+1$ can be expressed as the sum of two squares.

(i) Let $p = 4k+1$ be prime. Using the previous problem and Problem 5, prove that the period of $\sqrt{p}$ is odd and deduce that $\sqrt{p}$ has an expansion of the form

$$\sqrt{p} = \langle a_0; \overline{a_1, a_2, \ldots, a_{\ell-1}, a_\ell, a_\ell, a_{\ell-1}, \ldots, a_1, 2a_0}\rangle.$$

(ii) Let η be the complete quotient $\xi_{\ell+1}$:

$$\eta := \xi_{\ell+1} = \langle \overline{a_\ell; a_{\ell-1} \ldots, a_1, 2a_0, a_1, \ldots, a_{\ell-1}, a_\ell}\rangle.$$

Prove that $-1 = \eta \cdot \overline{\eta}$. Suggestion: Use Lemma 8.33.

(iii) Finally, writing $\eta = (a+\sqrt{p})/b$ (why does η have this form?) show that $p = a^2+b^2$.

8.10. Epilogue: Transcendental numbers, π, e, and where's calculus?

It's time to get a tissue box, because, unfortunately, our adventures through this book have come to an end. In this section we wrap up this book with a discussion on transcendental numbers and continued fractions.

8.10.1. Approximable numbers. A real number ξ is said to be **approximable** (by rationals) to order $n \geq 1$ if there exists a constant C and infinitely many rational numbers p/q in lowest terms with $q > 0$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{C}{q^n}. \tag{8.61}$$

Observe that if ξ is approximable to order $n > 1$, then it is automatically approximable to $n-1$; this is because

$$\left|\xi - \frac{p}{q}\right| < \frac{C}{q^n} \leq \frac{C}{q^{n-1}}.$$

Similarly, ξ approximable to any order k with $1 \leq k \leq n$. Intuitively, the approximability order n measures how close we can surround ξ with "good" rational numbers, that is, rational numbers having small denominators. To see what this means, suppose that ξ is only approximable to order 1. Thus, there is a C and infinitely many rational numbers p/q in lowest terms with $q > 0$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{C}{q}.$$

This inequality suggests that in order to find rational numbers very close to ξ, these rational numbers need to have large denominators to make C/q small. However, if ξ were approximable to order 1000, then there is a C and infinitely many rational numbers p/q in lowest terms with $q > 0$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{C}{q^{1000}}.$$

This inequality suggests that in order to find rational numbers very close to ξ, these rational numbers don't need to have large denominators, because even for small q, the large power of 1000 will make C/q^{1000} small. The following lemma shows that there is a limit to how close we can surround algebraic numbers by "good" rational numbers.

LEMMA 8.38. *If ξ is real algebraic of degree $n \geq 1$ (so ξ is rational if $n = 1$), then there exists a constant $c > 0$ such that for all rational numbers $p/q \neq \xi$ with $q > 0$, we have*

$$\left|\xi - \frac{p}{q}\right| \geq \frac{c}{q^n}.$$

PROOF. Assume that $f(\xi) = 0$ where

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0, \qquad a_k \in \mathbb{Z},$$

and that no such polynomial function of lower degree has this property. First, we claim that $f(r) \neq 0$ for any rational number $r \neq \xi$. Indeed, if $f(r) = 0$ for some rational number $r \neq \xi$, then we can write $f(x) = (x-r)g(x)$ where g is a polynomial of degree $n-1$. Then $0 = f(\xi) = (\xi - r)g(\xi)$ implies, since $\xi \neq r$, that $g(\xi) = 0$. This implies that the degree of ξ is $n-1$ contradicting the fact that the degree of ξ is n. Now for any rational $p/q \neq \xi$ with $q > 0$, we see that

$$\begin{aligned} 0 \neq |f(p/q)| &= \left|a_n\Big(\frac{p}{q}\Big)^n + a_{n-1}\Big(\frac{p}{q}\Big)^{n-1} + \cdots + a_1\Big(\frac{p}{q}\Big) + a_0\right| \\ &= \frac{|a_n p^n + a_{n-1}p^{n-1}q + \cdots + a_1 p q^{n-1} + a_0 q^n|}{q^n}. \end{aligned}$$

The numerator is a nonnegative integer, which cannot be zero, so the numerator must be ≥ 1. Therefore,

$$(8.62) \qquad |f(p/q)| \geq 1/q^n \ \text{ for all rational numbers } p/q \neq \xi \text{ with } q > 0.$$

Second, we claim that there is an $M > 0$ such that

$$(8.63) \qquad |x - \xi| \leq 1 \quad \Longrightarrow \quad |f(x)| \leq M|x - \xi|.$$

Indeed, note that since $f(\xi) = 0$, we have

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

Since

$$x^k - \xi^k = (x - \xi)\, q_k(x), \quad q_k(x) = x^{k-1} + x^{k-2}\,\xi + \cdots + x\,\xi^{k-2} + \xi^{k-1},$$

plugging each of these, for $k = 1, 2, 3, \ldots, n$, into the previous equation for $f(x)$, we see that $f(x) = (x - \xi)h(x)$ where h is a continuous function. In particular, since $[\xi - 1, \xi + 1]$ is a closed and bounded interval, there is an M such that $|h(x)| \leq M$ for all $x \in [\xi - 1, \xi + 1]$. This proves our claim.

Finally, let $p/q \neq \xi$ be a rational number with $q > 0$. If $|\xi - p/q| > 1$, then

$$\left|\xi - \frac{p}{q}\right| > 1 \geq \frac{1}{q^n}.$$

If $|\xi - p/q| \leq 1$, then by (8.62) and (8.63), we have

$$\left|\xi - \frac{p}{q}\right| \geq \frac{1}{M}|f(p/q)| \geq \frac{1}{M}\frac{1}{q^n}.$$

Hence, $|\xi - p/q| \geq c/q^n$ for all rational $p/q \neq \xi$ with $q > 0$, where c is the smaller of 1 and $1/M$. □

Let us form the contrapositive of the statement of this lemma: If $n \in \mathbb{N}$ and for all constants $c > 0$, there exists a rational number $p/q \neq \xi$ with $q > 0$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{c}{q^n}, \tag{8.64}$$

then ξ is not algebraic of degree n. Since a transcendental number is a number that is not algebraic of any degree n, we can think of a transcendental number as a number that can be surrounded arbitrarily close by "good" rational numbers. This leads us to Liouville numbers to be discussed shortly, but before talking about these special transcendental numbers, we use our lemma to prove the following important result.

Theorem 8.39. *A real algebraic number of degree n is not approximable to order $n+1$ (and hence not to any higher order). Moreover, a rational number is approximable to order 1 and a real number is irrational if and only if it is approximable to order 2.*

Proof. Let ξ be algebraic of degree $n \geq 1$ (so ξ is rational if $n = 1$). Then by Lemma 8.38, there exists a constant c such that for all rational numbers $p/q \neq \xi$ with $q > 0$, we have

$$\left|\xi - \frac{p}{q}\right| \geq \frac{c}{q^n}.$$

It follows that ξ is not approximable by rationals to order $n+1$ because

$$\left|\xi - \frac{p}{q}\right| < \frac{C}{q^{n+1}} \quad \Longrightarrow \quad \frac{c}{q^n} < \frac{C}{q^{n+1}} \quad \Longrightarrow \quad q < C/c.$$

Since there are only finitely many integers q such that $q < C/c$; it follows that there are only finitely many fractions p/q such that $|\xi - p/q| < C/q^{n+1}$.

Let a/b be a rational number in lowest terms with $b \geq 1$; we shall prove that a/b is approximable to order 1. (Note that we already know from our first statement that a/b is not approximable to order 2.) From Theorem 8.9, we know that the equation $ax - by = 1$ has an infinite number of integer solutions (x, y). The solutions (x, y) are automatically relatively prime. Moreover, if (x_0, y_0) is any one integral solution, then all solutions are of the form

$$x = x_0 + bt \quad , \quad y = y_0 + at \quad , \quad t \in \mathbb{Z}.$$

Since $b \geq 1$ we can choose t large so as to get infinitely many solutions with $x > 0$. With $x > 0$, we see that

$$\left|\frac{a}{b} - \frac{y}{x}\right| = \left|\frac{ax - by}{bx}\right| = \frac{1}{bx} < \frac{2}{x},$$

which shows that a/b is approximable to order 1.

Finally, if a number is irrational, then it is approximable to order 2 from Dirichlet's approximation theorem 8.21; conversely, if a number is approximable to order 2, then it must be irrational by the first statement of this theorem. □

Using this theorem we can prove that certain numbers must be irrational. For instance, let $\{a_n\}$ be any sequence of $0, 1$'s where there are infinitely many 1's. Consider

$$\xi = \sum_{n=0}^{\infty} \frac{a_n}{2^{2^n}}.$$

Note that ξ is the real number with binary expansion $a_0.0a_10a_20\cdots$, with a_n in the 2^n-th decimal place and with zeros everywhere else. Any case, fix a natural number n with $a_n \neq 0$ and let $s_n = \sum_{k=0}^{n} \frac{a_k}{2^{2^k}}$ be the n-th partial sum of this series. Then we can write s_n as p/q where $q = 2^{2^n}$. Observe that

$$\begin{aligned}\big|\xi - s_n\big| &\leq \frac{1}{2^{2^{n+1}}} + \frac{1}{2^{2^{n+2}}} + \frac{1}{2^{2^{n+3}}} + \frac{1}{2^{2^{n+4}}} + \cdots \\ &< \frac{1}{2^{2^{n+1}}} + \frac{1}{2^{2^{n+1}+1}} + \frac{1}{2^{2^{n+1}+2}} + \frac{1}{2^{2^{n+1}+3}} + \cdots \\ &= \frac{1}{2^{2^{n+1}}}\Big(1 + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots\Big) = \frac{2}{2^{2^{n+1}}} = \frac{2}{(2^{2^n})^2}.\end{aligned}$$

In conclusion,

$$\big|\xi - s_n\big| < \frac{2}{(2^{2^n})^2} = \frac{C}{q^2},$$

where $C = 2$. Thus, ξ is approximable to order 2, and hence must be irrational.

8.10.2. Liouville numbers. Numbers that satisfy (8.64) with $c = 1$ are special: A real number ξ is called a **Liouville number**, after Joseph Liouville (1809–1882), if for every natural number n there is a rational number $p/q \neq \xi$ with $q > 1$ such that

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{q^n}.$$

These numbers are transcendental by our discussion around (8.64). Because this fact is so important, we state this as a theorem.

THEOREM 8.40 (**Liouville's theorem**). *Any Liouville number is transcendental.*

Using Liouville's theorem we can give many (in fact uncountably many — see Problem 3) examples of transcendental numbers. Let $\{a_n\}$ be any sequence of integers in $0, 1, \ldots, 9$ where there are infinitely many nonzero integers. Let

$$\xi = \sum_{n=0}^{\infty} \frac{a_n}{10^{n!}}.$$

Note that ξ is the real number with decimal expansion

$$a_0.a_1a_2000a_3000000000000000000a_4\cdots,$$

with a_n in the $n!$-th decimal place and with zeros everywhere else. Using Liouville's theorem we'll show that ξ is transcendental. Fix a natural number n with $a_n \neq 0$ and let s_n be the n-th partial sum of this series. Then s_n can be written as p/q where $q = 10^{n!} > 1$. Observe that

$$\begin{aligned}\big|\xi - s_n\big| &\leq \frac{9}{10^{(n+1)!}} + \frac{9}{10^{(n+2)!}} + \frac{9}{10^{(n+3)!}} + \frac{9}{10^{(n+4)!}} + \cdots \\ &< \frac{9}{10^{(n+1)!}} + \frac{9}{10^{(n+1)!+1}} + \frac{9}{10^{(n+1)!+2}} + \frac{9}{10^{(n+1)!+3}} + \cdots \\ &= \frac{9}{10^{(n+1)!}}\Big(1 + \frac{1}{10^1} + \frac{1}{10^2} + \frac{1}{10^3} + \cdots\Big) \\ &= \frac{10}{10^{(n+1)!}} = \frac{10}{10^{n\cdot n!}\cdot 10^{n!}} \leq \frac{1}{10^{n\cdot n!}}.\end{aligned}$$

In conclusion,

$$\bigl|\xi - s_n\bigr| < \frac{1}{(10^{n!})^n} = \frac{1}{q^n},$$

so ξ is a Liouville number and therefore is transcendental.

8.10.3. Continued fractions and the "most extreme" irrational of all irrational numbers. We now show how continued fractions can be used to *construct* transcendental numbers! This is achieved by the following simple observation. Let $\xi = \langle a_0; a_1, \ldots\rangle$ be an irrational real number written as a simple continued fraction and let $\{p_n/q_n\}$ be its convergents. Then by our fundamental approximation theorem 8.18, we know that

$$\left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}}.$$

Since

$$q_n q_{n+1} = q_n(a_{n+1} q_n + q_{n-1}) \geq a_{n+1} q_n^2,$$

we see that

$$\left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{a_{n+1}\, q_n^2}. \tag{8.65}$$

Thus, we can make the rational number p_n/q_n approximate ξ as close as we wish by simply taking the next partial quotient a_{n+1} larger. We use this observation in the following theorem.

Theorem 8.41. *Let $\varphi : \mathbb{N} \to (0, \infty)$ be a function. Then there is an irrational number ξ and infinitely many rational numbers p/q such that*

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{\varphi(q)}.$$

Proof. We define $\xi = \langle a_0; a_1, a_2, \ldots\rangle$ by choosing the a_n's inductively as follows. Let $a_0 \in \mathbb{N}$ be arbitrary. Assume that $a_0, \ldots, a_n$ have been chosen. With q_n the denominator of $\langle a_0; a_1, \ldots, a_n\rangle$, choose (via Archimedean) $a_{n+1} \in \mathbb{N}$ such that

$$a_{n+1} q_n^2 > \varphi(q_n).$$

This defines $\{a_n\}$. Now defining $\xi := \langle a_0; a_1, a_2, \ldots\rangle$, by (8.65), for any natural number n we have

$$\left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{a_{n+1}\, q_n^2} < \frac{1}{\varphi(q_n)}.$$

This completes our proof. □

Using this theorem we can easily find transcendental numbers. For example, with $\varphi(q) = e^q$, we can find an irrational ξ such that for infinitely many rational numbers p/q, we have

$$\left|\xi - \frac{p}{q}\right| < \frac{1}{e^q}.$$

Since for any $n \in \mathbb{N}$, we have $e^q = \sum_{k=0}^\infty q^k/k! > q^n/n!$, it follows that for infinitely many rational numbers p/q, we have

$$\left|\xi - \frac{p}{q}\right| < \frac{\text{constant}}{q^n}.$$

In particular, ξ is transcendental.

As we have just seen, we can form transcendental numbers by choosing the partial quotients in an infinite simple continued fraction to be very large and transcendental numbers are the irrational numbers which are "closest" to good rational numbers. With this in mind, we can think of infinite continued fractions with small partial quotients as far from being transcendental or far from rational. Since 1 is the smallest natural number, we can consider the golden ratio

$$\Phi = \frac{1+\sqrt{5}}{2} = \langle 1; 1, 1, 1, 1, 1, 1, 1, \ldots \rangle$$

as being the "most extreme" or "most irrational" of all irrational numbers in the sense that it is the "farthest" irrational number from being transcendental or the "farthest" irrational number from being rational.

8.10.4. What about π and e and what about calculus? Above we have already seen examples (in fact, uncountably many — see Problem 3) of transcendental numbers and we even know how to construct them using continued fractions. However, these numbers seem in some sense to be "artificially" made. What about numbers that are more "natural" such as π and e? Are these numbers transcendental? In fact, these numbers do turn out to be transcendental, but the "easiest" proofs of these facts need the technology of calculus (derivatives) [**162, 163**]! Hopefully this might give one reason (amongst many others) to take more courses in analysis where the calculus is taught. **Advertisement ☺:** The book [**136**] is a sequel to the book you're holding and in it is the next adventure through topology and calculus and during our journey we'll prove that π and e are transcendental. However, if you choose to go on this adventure, we ask you to look back at all the amazing things that we've encountered during these past chapters — everything without using one single derivative or integral!

Exercises 8.10.

1. Given any integer $b \geq 2$, prove that $\xi = \sum_{n=0}^{\infty} b^{-2^n}$ is irrational.
2. Let $b \geq 2$ be an integer and let $\{a_n\}$ be any sequence of integers $0, 1, \ldots, b-1$ where there are infinitely many nonzero a_n's. Prove that $\xi = \sum_{n=1}^{\infty} a_n b^{-n!}$ is transcendental.
3. Using a Cantor diagonal argument as in the proof of Theorem 3.36, prove that the set of all numbers of the form $\xi = \sum_{n=0}^{\infty} \frac{a_n}{10^{n!}}$ where $a_n \in \{0, 1, 2, \ldots, 9\}$ is uncountable. That is, assume that the set of all such numbers is countable and construct a number of the same sort not in the set. Since we already showed that all these numbers are Liouville numbers, they are transcendental, so this argument provides another proof that the set of all transcendental numbers is uncountable.
4. Going through the construction of Theorem 8.41, define $\xi \in \mathbb{R}$ such that if $\{p_n/q_n\}$ are the convergents of its canonical continued fraction expansion, then for all n,
$$\left| \xi - \frac{p_n}{q_n} \right| < \frac{1}{q_n^n}.$$
Show that ξ is a Liouville number, and hence is transcendental.

Bibliography

1. Stephen Abbott, *Understanding analysis*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2001.
2. Aaron D. Abrams and Matteo J. Paris, *The probability that* $(a,b)=1$, The College Math. J. **23** (1992), no. 1, 47.
3. E. S. Allen, *The scientific work of Vito Volterra*, Amer. Math. Monthly **48** (1941), 516–519.
4. Nathan Altshiller and J.J. Ginsburg, *Solution to problem 460*, Amer. Math. Monthly **24** (1917), no. 1, 32–33.
5. Robert N. Andersen, Justin Stumpf, and Julie Tiller, *Let* π *be* 3, Math. Mag. **76** (2003), no. 3, 225–231.
6. Tom Apostol, *Another elementary proof of euler's formula for* $\zeta(2k)$, Amer. Math. Monthly **80** (1973), no. 4, 425–431.
7. Tom M. Apostol, *Mathematical analysis: a modern approach to advanced calculus*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1957.
8. ______, *Irrationality of the square root of two – a geometric proof*, Amer. Math. Monthly **107** (2000), no. 9, 841–842.
9. R.C. Archibald, *Mathematicians and music*, Amer. Math. Monthly **31** (1924), no. 1, 1–25.
10. Jörg Arndt and Christoph Haenel, *Pi—unleashed*, second ed., Springer-Verlag, Berlin, 2001, Translated from the 1998 German original by Catriona Lischka and David Lischka, With 1 CD-ROM (Windows).
11. Raymond Ayoub, *Euler and the zeta function*, Amer. Math. Monthly **81** (1974), no. 10, 1067–1086.
12. Bruce S. Babcock and John W. Dawson, Jr., *A neglected approach to the logarithm*, Two Year College Math. J. **9** (1978), no. 3, 136–140.
13. D. H. Bailey, J. M. Borwein, P. B. Borwein, and S. Plouffe, *The quest for pi*, Math. Intelligencer **19** (1997), no. 1, 50–57.
14. W.W. Ball, *Short account of the history of mathematics*, fourth ed., Dover Publications Inc., New York, 1960.
15. J.M. Barbour, *Music and ternary continued fractions*, Amer. Math. Monthly **55** (1948), no. 9, 545–555.
16. C.W. Barnes, *Euler's constant and e*, Amer. Math. Monthly **91** (1984), no. 7, 428–430.
17. Robert G. Bartle and Donald R. Sherbert, *Introduction to real analysis*, second ed., John Wiley & Sons Inc., New York, 1992.
18. A.F. Beardon, *Sums of powers of integers*, Amer. Math. Monthly **103** (1996), no. 3, 201–213.
19. A.H. Bell, *The "cattle problem." by Archimedies 251 b. c.*, Amer. Math. Monthly **2** (1885), no. 5, 140–141.
20. Howard E. Bell, *Proof of a fundamental theorem on sequences*, Amer. Math. Monthly **71** (1964), no. 6, 665–666.
21. Jordan Bell, *On the sums of series of reciprocals*, Available at `http://arxiv.org/abs/math/0506415`. Originally published as *De summis serierum reciprocarum*, Commentarii academiae scientiarum Petropolitanae 7 (1740) 123134 and reprinted in Leonhard Euler, Opera Omnia, Series 1: Opera mathematica, Volume 14, Birkhäuser, 1992. Original text, numbered E41, is available at the Euler Archive, `http://www.eulerarchive.org`.
22. W. W. Bell, *Special functions for scientists and engineers*, Dover Publications Inc., Mineola, NY, 2004, Reprint of the 1968 original.
23. Richard Bellman, *A note on the divergence of a series*, Amer. Math. Monthly **50** (1943), no. 5, 318–319.

24. Paul Benacerraf and Hilary Putnam (eds.), *Philosophy of mathematics: selected readings*, Cambridge University Press, Cambridge, 1964.
25. Stanley J. Benkoski, *The probability that k positive integers are relatively r-prime*, J. Number Theory **8** (1976), no. 2, 218–223.
26. Lennart Berggren, Jonathan Borwein, and Peter Borwein, *Pi: a source book*, third ed., Springer-Verlag, New York, 2004.
27. Bruce C. Berndt, *Ramanujan's notebooks*, Math. Mag. **51** (1978), no. 3, 147–164.
28. N.M. Beskin, *Fascinating fractions*, Mir Publishers, Moscow, 1980, Translated by V.I. Kisln, 1986.
29. F. Beukers, *A note on the irrationality of $\zeta(2)$ and $\zeta(3)$*, Bull. London Math. Soc. **11** (1979), no. 3, 268–272.
30. Ralph P. Boas, *A primer of real functions*, fourth ed., Carus Mathematical Monographs, vol. 13, Mathematical Association of America, Washington, DC, 1996, Revised and with a preface by Harold P. Boas.
31. R.P. Boas, *Tannery's theorem*, Math. Mag. **38** (1965), no. 2, 64–66.
32. J.M. Borwein and Borwein P.B., *Ramanujan, modular equations, and approximations to pi or how to compute one billion digits of pi*, Amer. Math. Monthly **96** (1989), no. 3, 201–219.
33. Jonathan M. Borwein and Peter B. Borwein, *Pi and the AGM*, Canadian Mathematical Society Series of Monographs and Advanced Texts, 4, John Wiley & Sons Inc., New York, 1998, A study in analytic number theory and computational complexity, Reprint of the 1987 original, A Wiley-Interscience Publication.
34. R.H.M. Bosanquet, *An elementary treatise on musical intervals and temperament (london, 1876)*, Diapason press, Utrecht, 1987.
35. Carl B. Boyer, *Fermat's integration of X^n*, Nat. Math. Mag. **20** (1945), 29–32.
36. ______, *A history of mathematics*, second ed., John Wiley & Sons Inc., New York, 1991, With a foreword by Isaac Asimov, Revised and with a preface by Uta C. Merzbach.
37. Paul Bracken and Bruce S. Burdick, *Euler's formula for zeta function convolutions: 10754*, Amer. Math. Monthly **108** (2001), no. 8, 771–773.
38. David Bressoud, *Was calculus invented in India?*, College Math. J. **33** (2002), no. 1, 2–13.
39. David Brewster, *Letters of Euler to a german princess on different subjects in physics and philosophy*, Harper and Brothers, New York, 1834, In two volumes.
40. W.E. Briggs and Nick Franceschine, *Problem 1302*, Math. Mag. **62** (1989), no. 4, 275–276.
41. T.J. I'A. Bromwich, *An introduction to the theory of infinite series*, second ed., Macmillan, London, 1926.
42. Richard A. Brualdi, *Mathematical notes*, Amer. Math. Monthly **84** (1977), no. 10, 803–807.
43. Robert Bumcrot, *Irrationality made easy*, The College Math. J. **17** (1986), no. 3, 243–244.
44. Frank Burk, *Euler's constant*, The College Math. J. **16** (1985), no. 4, 279.
45. Florian Cajori, *A history of mathematical notations*, Dover Publications Inc., New York, 1993, 2 Vol in 1 edition.
46. B.C. Carlson, *Algorithms involving arithmetic and geometric means*, Amer. Math. Monthly **78** (1971), 496–505.
47. Dario Castellanos, *The ubiquitous π*, Math. Mag. **61** (1988), no. 2, 67–98.
48. ______, *The ubiquitous π*, Math. Mag. **61** (1988), no. 3, 148–163.
49. R. Chapman, *Evaluating $\zeta(2)$*, preprint, 1999.
50. Robert R. Christian, *Another completeness property*, Amer. Math. Monthly **71** (1964), no. 1, 78.
51. James A. Clarkson, *On the series of prime reciprocals*, Proc. Amer. Math. Soc. **17** (1966), no. 2, 541.
52. Benoit Cloitre, private communication.
53. J. Brian Conrey, *The Riemann hypothesis*, Notices Amer. Math. Soc. **50** (2003), no. 3, 341–353.
54. F. Lee Cook, *A simple explicit formula for the Bernoulli numbers*, Two Year College Math. J. **13** (1982), no. 4, 273–274.
55. J. L. Coolidge, *The number e*, Amer. Math. Monthly **57** (1950), 591–602.
56. Fr. Gabe Costa, *Solution 277*, The College Math. J. **17** (1986), no. 1, 98–99.
57. Richard Courant and Herbert Robbins, *What is mathematics?*, Oxford University Press, New York, 1979, An elementary approach to ideas and methods.

58. E. J. Dijksterhuis, *Archimedes*, Princeton University Press, Princeton, NJ, 1987, Translated from the Dutch by C. Dikshoorn, Reprint of the 1956 edition, With a contribution by Wilbur R. Knorr.
59. Underwood Dudley, *A budget of trisections*, Springer-Verlag, New York, 1987.
60. William Dunham, *A historical gem from Vito Volterra*, Math. Mag. **63** (1990), no. 4, 234–237.
61. ———, *Euler and the fundamental theorem of algebra*, The College Math. J. **22** (1991), no. 4, 282–293.
62. E. Dunne and M. Mcconnell, *Pianos and continued fractions*, Math. Mag. **72** (1999), no. 2, 104–115.
63. Erich Dux, *Ein kurzer Beweis der Divergenz der unendlichen Reihe* $\sum_{r=1}^{\infty} 1/p_r$, Elem. Math. **11** (1956), 50–51.
64. Erdös, *Uber die Reihe* $\sum 1/p$, Mathematica Zutphen. B. **7** (1938), 1–2.
65. Leonhard Euler, *Introduction to analysis of the infinite. Book I*, Springer-Verlag, New York, 1988, Translated from the Latin and with an introduction by John D. Blanton.
66. ———, *Introduction to analysis of the infinite. Book II*, Springer-Verlag, New York, 1990, Translated from the Latin and with an introduction by John D. Blanton.
67. H Eves, *Mathematical circles squared*, Prindle Weber & Schmidt, Boston, 1972.
68. Pierre Eymard and Jean-Pierre Lafon, *The number* π, American Mathematical Society, Providence, RI, 2004, Translated from the 1999 French original by Stephen S. Wilson.
69. Charles Fefferman, *An easy proof of the fundmental theorem of algebra*, Amer. Math. Monthly **74** (1967), no. 7, 854–855.
70. William Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons Inc., New York, 1968.
71. ———, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971.
72. D. Ferguson, *Evaluation of* π*. are shanks' figures correct?*, Mathematical Gazette **30** (1946), 89–90.
73. William Leonard Ferrar, *A textbook of convergence*, The Clarendon Press Oxford University Press, New York, 1980.
74. Steven R. Finch, *Mathematical constants*, Encyclopedia of Mathematics and its Applications, vol. 94, Cambridge University Press, Cambridge, 2003.
75. Philippe Flajolet and Ilan Vardi, *Zeta function expansions of classical constants*, preprint, 1996.
76. Tomlinson Fort, *Application of the summation by parts formula to summability of series*, Math. Mag. **26** (1953), no. 26, 199–204.
77. Gregory Fredricks and Roger B. Nelsen, *Summation by parts*, The College Math. J. **23** (1992), no. 1, 39–42.
78. Richard J. Friedlander, *Factoring factorials*, Two Year College Math. J. **12** (1981), no. 1, 12–20.
79. Joseph A. Gallian, *contemporary abstract algebra*, sixth ed., Houghton Mifflin Company, Boston, 2005.
80. Martin Gardner, *Mathematical games*, Scientific American **April** (1958).
81. ———, *Second scientific american book of mathematical puzzles and diversions*, University of Chicago press, Chicago, 1987, Reprint edition.
82. J. Glaisher, *History of Euler's constant*, Messenger of Math. **1** (1872), 25–30.
83. Edward J. Goodwin, *Quadrature of the circle*, Amer. Math. Monthly **1** (1894), no. 1, 246–247.
84. Russell A. Gordon, *The use of tagged partitions in elementary real analysis*, Amer. Math. Monthly **105** (1998), no. 2, 107–117.
85. H.W. Gould, *Explicit formulas for Bernoulli numbers*, Amer. Math. Monthly **79** (1972), no. 1, 44–51.
86. D.S. Greenstein, *A property of the logarithm*, Amer. Math. Monthly **72** (1965), no. 7, 767.
87. Robert Grey, *Georg Cantor and transcendental numbers*, Amer. Math. Monthly **101** (1994), no. 9, 819–832.
88. Lucye Guilbeau, *The history of the solution of the cubic equation*, Mathematics News Letter **5** (1930), no. 4, 8–12.

89. Rachel W. Hall and Krešimir Josić, *The mathematics of musical instruments*, Amer. Math. Monthly **108** (2001), no. 4, 347–357.
90. Hallerberg, *Indiana's squared circle*, Math. Mag. **50** (1977), no. 3, 136–140.
91. Paul R. Halmos, *Naive set theory*, Springer-Verlag, New York-Heidelberg, 1974, Reprint of the 1960 edition. Undergraduate Texts in Mathematics.
92. ______, *I want to be a mathematician*, Springer-Verlag, 1985, An automathography.
93. G.D. Halsey and Edwin Hewitt, *More on the superparticular ratios in music*, Amer. Math. Monthly **79** (1972), no. 10, 1096–1100.
94. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1988, Reprint of the 1952 edition.
95. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, fifth ed., The Clarendon Press Oxford University Press, New York, 1979.
96. Julian Havil, *Gamma*, Princeton University Press, Princeton, NJ, 2003, Exploring Euler's constant, With a foreward by Freeman Dyson.
97. Ko Hayashi, *Fibonacci numbers and the arctangent function*, Math. Mag. **76** (2003), no. 3, 214–215.
98. T. L. Heath, *Diophantus of alexandria: a study in the history of greek algebra*, Cambridge University Press, England, 1889.
99. ______, *The works of Archimedes*, Cambridge University Press, England, 1897.
100. Thomas Heath, *A history of Greek mathematics. Vol. I*, Dover Publications Inc., New York, 1981, From Thales to Euclid, Corrected reprint of the 1921 original.
101. Aaron Herschfeld, *On Infinite Radicals*, Amer. Math. Monthly **42** (1935), no. 7, 419–429.
102. Josef Hofbauer, *A simple proof of* $1+1/2^2+1/3^2+\cdots=\pi^2/6$ *and related identities*, Amer. Math. Monthly **109** (2002), no. 2, 196–200.
103. P. Iain, *Science, theology and einstein*, Oxford University, Oxford, 1982.
104. Frank Irwin, *A curious convergent series*, Amer. Math. Monthly **23** (1916), no. 5, 149–152.
105. Sir James H. Jeans, *Science and music*, Dover Publications Inc., New York, 1968, Reprint of the 1937 edition.
106. Dixon J. Jones, *Continued powers and a sufficient condition for their convergence*, Math. Mag. **68** (1995), no. 5, 387–392.
107. Gareth A. Jones, $6/\pi^2$, Math. Mag. **66** (1993), no. 5, 290–298.
108. J.P. Jones and S. Toporowski, *Irrational numbers*, Amer. Math. Monthly **80** (1973), no. 4, 423–424.
109. Dan Kalman, *Six ways to sum a series*, The College Math. J. **24** (1993), no. 5, 402–421.
110. Edward Kasner and James Newman, *Mathematics and the imagination*, Dover Publications Inc., New York, 2001.
111. Victor J. Katz, *Ideas of calculus in islam and india*, Math. Mag. **68** (1995), no. 3, 163–174.
112. Gerard W. Kelly, *Short-cut math*, Dover Publications Inc., New York, 1984.
113. A. J. Kempner, *A curious convergent series*, Amer. Math. Monthly **21** (1914), no. 2, 48–50.
114. Alexey Nikolaevitch Khovanskii, *The application of continued fractions and their generalizations to problems in approximation theory*, Translated by Peter Wynn, P. Noordhoff N. V., Groningen, 1963.
115. Steven J. Kifowit and Terra A. Stamps, *The harmonic series diverges again and again*, The AMATYC Review **27** (2006), no. 2, 31–43.
116. M.S. Klamkin and Robert Steinberg, *Problem 4431*, Amer. Math. Monthly **59** (1952), no. 7, 471–472.
117. M.S. Klamkin and J.V. Whittaker, *Problem 4564*, Amer. Math. Monthly **62** (1955), no. 2, 129–130.
118. Israel Kleiner, *Evolution of the function concept: A brief survey*, Two Year College Math. J. **20** (1989), no. 4, 282–300.
119. Morris Kline, *Euler and infinite series*, Math. Mag. **56** (1983), no. 5, 307–314.
120. Konrad Knopp, *Infinite sequences and series*, Dover Publications Inc., New York, 1956, Translated by Frederick Bagemihl.
121. R. Knott, *Fibonacci numbers and the golden section*, `http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/` .
122. Donald E. Knuth, *The art of computer programming. Vol. 2*, second ed., Addison-Wesley Publishing Co., Reading, Mass., 1981, Seminumerical algorithms, Addison-Wesley Series in Computer Science and Information Processing.

123. R.A. Kortram, *Simple proofs for* $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$ *and* $\sin x = x \prod_{k=1}^{\infty}(1 - x^2/k^2\pi^2)$, Math. Mag. **69** (1996), no. 2, 122–125.

124. Myren Krom, *On sums of powers of natural numbers*, Two Year College Math. J. **14** (1983), no. 4, 349–351.

125. David E. Kullman, *What's harmonic about the harmonic series*, The College Math. J. **32** (2001), no. 3, 201–203.

126. R. Kumanduri and C. Romero, *Number theory with computer applications*, Prentice-Hall, Simon and Schuster, New Jersey, 1998.

127. M. Laczkovich, *On Lambert's proof of the irrationality of* π, Amer. Math. Monthly **104** (1997), no. 5, 439–443.

128. Serge Lang, *A first course in calculus*, fifth ed., Addison-Wesley Pub. Co., Reading, Mass., 1964.

129. L. J. Lange, *An elegant continued fraction for* π, Amer. Math. Monthly **106** (1999), no. 5, 456–458.

130. W. G. Leavitt, *The sum of the reciprocals of the primes*, Two Year College Math. J. **10** (1979), no. 3, 198–199.

131. D.H. Lehmer, *Problem 3801*, Amer. Math. Monthly **43** (1936), no. 9, 580.

132. ———, *On arccotangent relations for* π, Amer. Math. Monthly **45** (1938), no. 10, 657–664.

133. D.H. Lehmer and M.A. Heaslet, *Solution 3801*, Amer. Math. Monthly **45** (1938), no. 9, 636–637.

134. A.L. Leigh Silver, *Musimatics or the nun's fiddle*, Amer. Math. Monthly **78** (1971), no. 4, 351–357.

135. H.W. Lenstra, *Solving the pell equation*, Notices Amer. Math. Soc. **49** (2002), no. 2, 182–192.

136. P. Loya, *Amazing and aesthetic aspects of analysis: The celebrated calculus*, in preparation.

137. N. Luzin, *Function: Part I*, Amer. Math. Monthly **105** (1998), no. 1, 59–67.

138. ———, *Function: Part II*, Amer. Math. Monthly **105** (1998), no. 3, 263–270.

139. Richard Lyon and Morgan Ward, *The limit for e*, Amer. Math. Monthly **59** (1952), no. 2, 102–103.

140. Desmond MacHales, *Comic sections: The book of mathematical jokes, humour, wit, and wisdom*, Boole Press, Dublin, 1993.

141. Alan L. Mackay, *Dictionary of scientific quotations*, Institute of Physics Publishing, Bristol, 1994.

142. E.A. Maier, *On the irrationality of certain trigonometric numbers*, Amer. Math. Monthly **72** (1965), no. 9, 1012–1013.

143. E.A. Maier and Ivan Niven, *A method of establishing certain irrationalities*, Math. Mag. **37** (1964), no. 4, 208–210.

144. S. C. Malik, *Introduction to convergence*, Halsted Press, a division of John Wiley and sons, New Delhi, 1984.

145. Eli Maor, *e: the story of a number*, Princeton University Press, Princetown, NJ, 1994.

146. George Markowsky, *Misconceptions about the golden ratio*, Two Year College Math. J. **23** (1992), no. 1, 2–19.

147. Jerold Mathews, *Gear trains and continued fractions*, Amer. Math. Monthly **97** (1990), no. 6, 505–510.

148. Marcin Mazur, *Irrationality of* $\sqrt{2}$, private communication, 2004.

149. J. H. McKay, *The william lowell putnam mathematical competition*, Amer. Math. Monthly **74** (1967), no. 7, 771–777.

150. George Miel, *Of calculations past and present: The Archimedean algorithm*, Amer. Math. Monthly **90** (1983), no. 1, 17–35.

151. Jeff Miller, *Earliest uses of symbols in probability and statistics*, `http://members.aol.com/jeff570/stat.html`.

152. John E. Morrill, *Set theory and the indicator function*, Amer. Math. Monthly **89** (1982), no. 9, 694–695.

153. Leo Moser, *On the series,* $\sum 1/p$, Amer. Math. Monthly **65** (1958), 104–105.

154. Joseph Amal Nathan, *The irrationality of* e^x *for nonzero rational* x, Amer. Math. Monthly **105** (1998), no. 8, 762–763.

155. Harry L. Nelson, *A solution to Archimedes' cattle problem*, J. Recreational Math. **13** (1980-81), 162–176.

156. D.J. Newman, *Solution to problem e924*, Amer. Math. Monthly **58** (1951), no. 3, 190–191.

157. ———, *Arithmetic, geometric inequality*, Amer. Math. Monthly **67** (1960), no. 9, 886.
158. Donald J. Newman and T.D. Parsons, *On monotone subsequences*, Amer. Math. Monthly **95** (1988), no. 1, 44–45.
159. James R. Newman (ed.), *The world of mathematics. Vol. 1*, Dover Publications Inc., Mineola, NY, 2000, Reprint of the 1956 original.
160. J.R. Newman (ed.), *The world of mathematics*, Simon and Schuster, New York, 1956.
161. James Nickel, *Mathematics: Is God silent?*, Ross House Books, Vallecito, California, 2001.
162. Ivan Niven, *The transcendence of* π, Amer. Math. Monthly **46** (1939), no. 8, 469–471.
163. ———, *Irrational numbers*, The Carus Mathematical Monographs, No. 11, The Mathematical Association of America. Distributed by John Wiley and Sons, Inc., New York, N.Y., 1956.
164. ———, *A proof of the divergence of* $\sum 1/p$, Amer. Math. Monthly **78** (1971), no. 3, 272–273.
165. Ivan Niven and Herbert S. Zuckerman, *An introduction to the theory of numbers*, third ed., John Wiley & Sons, Inc., New York-London-Sydney, 1972.
166. Jeffrey Nunemacher and Robert M. Young, *On the sum of consecutive kth powers*, Math. Mag. **60** (1987), no. 4, 237–238.
167. Mícheál Ó Searcóid, *Elements of abstract analysis*, Springer Undergraduate Mathematics Series, Springer-Verlag London Ltd., London, 2002.
168. University of St. Andrews, *A chronology of pi*,
`http://www-gap.dcs.st-and.ac.uk/˜history/HistTopics/Pi_chronology.html`.
169. ———, *Eudoxus of cnidus*,
`http://www-groups.dcs.st-and.ac.uk/h̃istory/Biographies/Eudoxus.html`.
170. ———, *A history of pi*,
`http://www-gap.dcs.st-and.ac.uk/˜history/HistTopics/Pi_through_the_ages.html`.
171. ———, *Leonhard Euler*,
`http://www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Euler.html`.
172. ———, *Madhava of sangamagramma*,
`http://www-gap.dcs.st-and.ac.uk/ history/Mathematicians/Madhava.html`.
173. C. D. Olds, *The simple continued fraction expansion of e*, Amer. Math. Monthly **77** (1970), no. 9, 968–974.
174. Geo. A. Osborne, *A problem in number theory*, Amer. Math. Monthly **21** (1914), no. 5, 148–150.
175. Thomas J. Osler, *The union of Vieta's and Wallis's products for pi*, Amer. Math. Monthly **106** (1999), no. 8, 774–776.
176. Thomas J. Osler and James Smoak, *A magic trick from fibonacci*, The College Math. J. **34** (2003), 58–60.
177. Thomas J. Osler and Nicholas Stugard, *A collection of numbers whose proof of irrationality is like that of the number e*, Math. Comput. Ed. **40** (2006), 103–107.
178. Thomas J. Osler and Michael Wilhelm, *Variations on Vieta's and Wallis's products for pi*, Math. Comput. Ed. **35** (2001), 225–232.
179. Ioannis Papadimitriou, *A simple proof of the formula* $\sum_{k=1}^{\infty} k^{-2} = \pi^2/6$, Amer. Math. Monthly **80** (1973), no. 4, 424–425.
180. L. L. Pennisi, *Elementary proof that e is irrational*, Amer. Math. Monthly **60** (1953), 474.
181. G.M. Phillips, *Archimedes the numerical analyst*, Amer. Math. Monthly **88** (1981), no. 3, 165–169.
182. R.C. Pierce, Jr., *A brief history of logarithms*, Two Year College Math. J. **8** (1977), no. 1, 22–26.
183. Alfred S. Posamentier and Ingmar Lehmann, π*: A biography of the world's most mysterious number*, Prometheus Books, Amherst, NY, 2004, With an afterword by Herbert A. Hauptman.
184. G. Baley Price, *Telescoping sums and the summation of sequences*, Two Year College Math. J. **4** (1973), no. 4, 16–29.
185. Raymond Redheffer, *What! another note just on the fundamental theorem of algebra*, Amer. Math. Monthly **71** (1964), no. 2, 180–185.
186. Reinhold Remmert, *Vom Fundamentalsatz der Algebra zum Satz von Gelfand-Mazur*, Math. Semesterber. **40** (1993), no. 1, 63–71.
187. Dorothy Rice, *History of* π *(or pi)*, Mathematics News Letter **2** (1928), 6–8.
188. N. Rose, *Mathematical maxims and minims*, Rome Press Inc., Raleigh, NC, 1988.

189. Tony Rothman, *Genius and biographers: The fictionalization of Evariste Galois*, Amer. Math. Monthly **89** (1982), no. 2, 84–106.

190. Ranjan Roy, *The discovery of the series formula for π by Leibniz, Gregory and Nilakantha*, Math. Mag. **63** (1990), no. 5, 291–306.

191. Walter Rudin, *Principles of mathematical analysis*, third ed., McGraw-Hill Book Co., New York, 1976, International Series in Pure and Applied Mathematics.

192. ______, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.

193. Oliver Sacks, *The man who mistook his wife for a hat : And other clinical tales*, Touchstone, New York, 1985.

194. Yoram Sagher, *Notes: What Pythagoras Could Have Done*, Amer. Math. Monthly **95** (1988), no. 2, 117.

195. E. Sandifer, *How euler did it*,
`http://www.maa.org/news/howeulerdidit.html`.

196. Norman Schaumberger, *An instant proof of $e^\pi > \pi^e$*, The College Math. J. **16** (1985), no. 4, 280.

197. Murray Schechter, *Tempered scales and continued fractions*, Amer. Math. Monthly **87** (1980), no. 1, 40–42.

198. Herman C. Schepler, *A chronology of pi*, Math. Mag. **23** (1950), no. 3, 165–170.

199. ______, *A chronology of pi*, Math. Mag. **23** (1950), no. 4, 216–228.

200. ______, *A chronology of pi*, Math. Mag. **23** (1950), no. 5, 279–283.

201. P.J. Schillo, *On primitive pythagorean triangles*, Amer. Math. Monthly **58** (1951), no. 1, 30–32.

202. Fred Schuh, *The master book of mathematical recreations*, Dover Publications Inc., New York, 1968, Translated by F. Göbel.

203. P. Sebah and X. Gourdon, *A collection of formulae for the Euler constant*,
`http://numbers.computation.free.fr/Constants/Gamma/gammaFormulas.pdf`.

204. ______, *A collection of series for π*,
`http://numbers.computation.free.fr/Constants/Pi/piSeries.html`.

205. ______, *The constant e and its computation*,
`http://numbers.computation.free.fr/Constants/constants.html`.

206. ______, *Introduction on Bernoulli's numbers*,
`http://numbers.computation.free.fr/Constants/constants.html`.

207. ______, *π and its computation through the ages*,
`http://numbers.computation.free.fr/Constants/constants.html`.

208. Allen A. Shaw, *Note on roman numerals*, Nat. Math. Mag. **13** (1938), no. 3, 127–128.

209. Georgi E. Shilov, *Elementary real and complex analysis*, english ed., Dover Publications Inc., Mineola, NY, 1996, Revised English edition translated from the Russian and edited by Richard A. Silverman.

210. G. F. Simmons, *Calculus gems*, Mcgraw Hill, Inc., New York, 1992.

211. J.G. Simmons, *A new look at an old function, $e^{i\theta}$*, The College Math. J. **26** (1995), no. 1, 6–10.

212. Sahib Singh, *On dividing coconuts: A linear diophantine problem*, The College Math. J. **28** (1997), no. 3, 203–204.

213. David Singmaster, *The legal values of pi*, Math. Intelligencer **7** (1985), no. 2, 69–72.

214. ______, *Coconuts: the history and solutions of a classic Diophantine problem*, Gaṇita-Bhāratī **19** (1997), no. 1-4, 35–51.

215. Walter S. Sizer, *Continued roots*, Math. Mag. **59** (1986), no. 1, 23–27.

216. David Eugene Smith, *A source book in mathematics. vol. 1, 2.*, Dover Publications, Inc, New York, 1959, Unabridged and unaltered republ. of the first ed. 1929.

217. J. Sondow, *Problem 88*, Math Horizons (1997), 32, 34.

218. H. Steinhaus, *Mathematical snapshots*, english ed., Dover Publications Inc., Mineola, NY, 1999, Translated from the Polish, With a preface by Morris Kline.

219. Ian Stewart, *Concepts of modern mathematics*, Dover Publications Inc., New York, 1995.

220. John Stillwell, *Galois theory for beginners*, Amer. Math. Monthly **101** (1994), no. 1, 22–27.

221. D. J. Struik (ed.), *A source book in mathematics, 1200–1800*, Princeton Paperbacks, Princeton University Press, Princeton, NJ, 1986, Reprint of the 1969 edition.

222. Frode Terkelsen, *The fundamental theorem of algebra*, Amer. Math. Monthly **83** (1976), no. 8, 647.

223. Hugh Thurston, *A simple proof that every sequence has a monotone subsequence*, Math. Mag. **67** (1994), no. 5, 344.
224. C. Tøndering, *Frequently asked questions about calendars*, `http://www.tondering.dk/claus/`, 2003.
225. Herbert Turnbull, *The great mathematicians*, Barnes & Noble, New York, 1993.
226. Herbert (ed.) Turnbull, *The correspondence of Isaac Newton, Vol. II: 1676–1687*, Published for the Royal Society, Cambridge University Press, New York, 1960.
227. D. J. Uherka and Ann M. Sergott, *On the continuous dependence of the roots of a polynomial on its coefficients*, Amer. Math. Monthly **84** (1977), no. 5, 368–370.
228. R.S. Underwood and Robert E. Moritz, *Solution to problem 3242*, Amer. Math. Monthly **35** (1928), no. 1, 47–48.
229. James Victor Uspensky, *Introduction to mathematical probability*, McGraw-Hill Book Co, New York, London, 1937.
230. Alfred van der Poorten, *A proof that Euler missed...Apéry's proof of the irrationality of* $\zeta(3)$, Math. Intelligencer **1** (1978/79), no. 4, 195–203, An informal report.
231. Charles Vanden Eynden, *Proofs that* $\sum 1/p$ *diverges*, Amer. Math. Monthly **87** (1980), no. 5, 394–397.
232. Ilan Vardi, *Computational recreations in Mathematica*, Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1991.
233. ———, *Archimedes' cattle problem*, Amer. Math. Monthly **105** (1998), no. 4, 305–319.
234. P.G.J. Vredenduin, *A paradox of set theory*, Amer. Math. Monthly **76** (1969), no. 1, 59–60.
235. A.D. Wadhwa, *An interesting subseries of the harmonic series*, Amer. Math. Monthly **82** (1975), no. 9, 931–933.
236. Morgan Ward, *A mnemonic for Euler's constant*, Amer. Math. Monthly **38** (1931), no. 9, 6.
237. André Weil, *Number theory*, Birkhäuser Boston Inc., Boston, MA, 1984, An approach through history, From Hammurapi to Legendre.
238. E. Weisstein, *Dirichlet function. from* MathWorld—*a wolfram web resource*, `http://mathworld.wolfram.com/DirichletFunction.html`.
239. ———, *Landau symbols. from* MathWorld—*a wolfram web resource*, `http://mathworld.wolfram.com/LandauSymbols.html`.
240. ———, *Pi approximations. from* MathWorld—*a wolfram web resource*, `http://mathworld.wolfram.com/PiApproximations.html`.
241. ———, *Pi formulas. from* MathWorld—*a wolfram web resource*, `http://mathworld.wolfram.com/PiFormulas.html`.
242. B. R. Wenner, *Continuous, exactly* k*-to-one functions on* $\mathbf{R}$, Math. Mag. **45** (1972), 224–225.
243. Joseph Wiener, *Bernoulli's inequality and the number* e, The College Math. J. **16** (1985), no. 5, 399–400.
244. E. Wigner, *The unreasonable effectiveness of mathematics in the natural sciences*, Comm. Pure Appl. Math. **13** (1960), 1–14.
245. Eugene Wigner, *Symmetries and reflections: Scientific essays*, The MIT press, Cambridge and London, 1970.
246. Herbert S. Wilf, *generatingfunctionology*, third ed., A K Peters Ltd., Wellesley, MA, 2006, Freely downloadable at `http://www.cis.upenn.edu/ wilf/`.
247. G.T. Williams, *A new method of evaluating* $\zeta(2n)$, Amer. Math. Monthly **60** (1953), no. 1, 12–25.
248. H.C. Williams, R.A. German, and C.R. Zarnke, *Solution of the cattle problem of Archimedes*, Math. Comp. **19** (1965), no. 92, 671–674.
249. A. M. Yaglom and I. M. Yaglom, *Challenging mathematical problems with elementary solutions. Vol. II*, Dover Publications Inc., New York, 1987, Problems from various branches of mathematics, Translated from the Russian by James McCawley, Jr., Reprint of the 1967 edition.
250. Hansheng Yang and Yang Heng, *The arithmetic-geometric mean inequality and the constant* e, Math. Mag. **74** (2001), no. 4, 321–323.
251. G.S. Young, *The linear functional equation*, Amer. Math. Monthly **65** (1958), no. 1, 37–38.
252. Robert M. Young, *Excursions in calculus*, The Dolciani Mathematical Expositions, vol. 13, Mathematical Association of America, Washington, DC, 1992, An interplay of the continuous and the discrete.
253. Don Zagier, *The first* 50 *million prime numbers*, Math. Intelligencer **0** (1977/78), 7–19.

254. Lee Zia, *Using the finite difference calculus to sum powers of integers*, The College Math. J. **22** (1991), no. 4, 294–300.

Index

Abel summable, 299
Abel's lemma, 271
Abel's limit theorem, 296
Abel's multiplication theorem, 317
Abel's test for series, 277
Abel, Neils, 296
Abel, Niels, 123, 271
Absolute convergence
  double series, 302
  infinite products, 355
  series, 135
Absolute convergence theorem, 135
Absolute value, 17, 39
  of complex numbers, 82
Absolute value rules, 39
Additive function, 171
Additive identity
  existence for complex numbers, 80
  existence for integers, 35
  existence for reals, 54
  existence for vectors, 73
Additive inverse
  existence for complex numbers, 80
  existence for integers, 35
  existence for reals, 54
  existence for vectors, 73
AGM method (to compute π), 233
Ahmes, 227
Algebra of limits
  functions, 161
  sequences, 107
Algebraic number, 88
Algorithm
  Archimedes', 232
  Borchardt's, 235
  Brahmagupta's, 461
  canonical continued fraction, 415
  division, 42
  Euclidean, 46
Almost integer, 206
Alternating
  harmonic series, 136, 154, 275
  harmonic series rearrangement, 311
  $\log 2$ formula, 194
  series error estimate, 275
  series test, 275
Angle, 78, 206
Anthoniszoon, Adriaan, 228, 423
Apéry, Roger, 263
Approximable numbers, 464
Approximation(s)
  best, 424
  good, 423
  $\pi \approx \sqrt{10}, \left(\frac{227}{23}\right)^{1/2}, 31^{1/3}, \left(\frac{4930}{159}\right)^{1/3}, 306^{1/5}, \left(\frac{77729}{254}\right)^{1/5}$, 432
  $\pi \approx \frac{6}{5}\Phi^2$, 414
  $\pi \approx \left(\frac{2143}{22}\right)^{1/4}$, 432
Approximations
  $e^\pi - \pi \approx 20$, 206
  $e^{\pi\sqrt{163}} = 262537412640768744$, 206
Archimedean
  ordering of the natural numbers, 25
  property of reals, 68
Archimedes of Syracuse, 153, 226, 227, 423, 456
Archimedes'
  algorithm, 232
Archimedes' cattle problem, 456
Archimedes' three propositions, 228
Argument
  of a complex number, 219
  principal, 219
Ariel, 437
Aristotle, 182
Aristoxenus, 437
Arithmetic mean, 34
Arithmetic properties of series theorem, 127
Arithmetic-geometric mean inequality, 34, 195
  application to e, 116
Associative law
  addition for complex numbers, 80
  addition for integers, 35
  addition for natural numbers, 22
  addition for reals, 53
  addition for vectors, 73
  multiplication for complex numbers, 80

multiplication for integers, 36
multiplication for natural numbers, 22
multiplication for reals, 54
multiplication for vectors, 73
Axiom(s)
completeness, 63
integers, 35
natural numbers, 22

b-adic expansion, *see also* b-adic expansion
b-adic representation
of integers, 51
of real numbers, 147
Bachmann, Paul, 291
Ball
closed, 77
open, 77
Ball norm, 76
Barber puzzle, 3
Base
of a power, 28, 190, 222
to represent integers, 51
to represent real numbers, 146
Basel problem, 250
Bell, Jordan, 238
Bellman, Richard, 321
Bernoulli numbers, 262, 270, 327
Bernoulli's inequality, 30
Bernoulli, Jacob, 30, 125, 249, 269
numbers, 270, 327
Bernoulli, Nicolaus, 344
Bertrand, Joseph, 293
Best approximation, 424
Best approximation theorem, 425, 429
big O notation, 291
Bijection, 18
Binary system, 50
Binomial
coefficient, 31, 332
series, 333, 336
theorem, 31
Bisection method, 181
Bolzano, Bernard, 114
Bolzano-Weierstrass theorem for $\mathbb{R}$, 114
Bolzano-Weierstrass theorem for $\mathbb{R}^m$, 114
Bombelli, Rafael, 414
Borchardt's algorithm, 235
Borchardt, Carl, 235
Borwein, Jonathan, 233
Borwein, Peter, 233
Bosanquet, Robert, 437
Bounded
above, 63
below, 64
sequence, 104
variation, 272
Boundedness theorem, 175
Box norm, 79
Brahmagupta, 459
quote, 459
zero and negative numbers, 21
Brahmagupta's algorithm, 461
Brent, Richard, 233
Bromwich, Thomas, 257
Brouncker, Lord William, 233, 389
Brouwer's fixed point theorem, 179
Brouwer, L E J, 179
Bunyakovskiĭ, Viktor, 74

Caesar, Julius, 433
Calendar
Gregorian, 433
Julian, 433
Persian, 437
Canonical (simple) continued fraction, 415
Canonical continued fraction algorithm, 415
Cantor's diagonal argument, 151
original version, 152
Cantor's first proof (of the uncountability of $\mathbb{R}$), 86
Cantor's second proof (of the uncountability of $\mathbb{R}$), 151
Cantor's theorem, 90
Cantor, Georg, 83
uncountability of transcendental numbers, 89
Cardinality, 83
Cartesian product, 15
Cassini's identity, 411
Castellanos, Dario, 342
Cataldi, Pietro, 421
Cattle problem, 456
Cauchy condensation test, 133
Cauchy criterion theorem, 119
Cauchy product, 315
Cauchy sequence, 117
Cauchy's arithmetic mean theorem, 344
Cauchy's criterion theorem for series, 132
Cauchy's double series theorem, 145, 305
Cauchy's multiplication theorem, 318
Cauchy's root test, 284
Cauchy, Augustin, 74, 315
Cauchy-Bunyakovskiĭ-Schwarz inequality, 74
Cauchy-Hadamard theorem, 287
Cauchy-Schwarz inequality, 74
Chain of intervals, 181
Change of base formula, 196
Characteristic function, 20
Chongzhi, Zu, *see also* Chung-Chi, Tsu
Chung-Chi, Tsu, 227, 423
Cloitre, Benoit, 386
Closed
ball, 77
interval, 4
natural number operations, 22
Coconut puzzles, 409

Codomain of function, 16
Coin game, 34
Commutative law
  addition for complex numbers, 80
  addition for integers, 35
  addition for natural numbers, 22
  addition for reals, 53
  addition for vectors, 73
  multiplication for complex numbers, 80
  multiplication for integers, 36
  multiplication for natural numbers, 22
  multiplication for reals, 54
Commutative ring, 36
Compact, 173
Compactness lemma, 173
Comparison test, 132
Complement of sets, 8
Complete quotients, 416
Completeness axiom, 63
Completeness property of $\mathbb{R}$, 64
Completeness property of $\mathbb{R}^m$, 120
Complex
  logarithm, 221
  power, 190, 222
Complex conjugate, 82
Complex numbers ($\mathbb{C}$), 79
Component functions, 161
Component theorem
  for continuity, 168
  for functions, 161
  for sequences, 103
Composite, 44
Composition of functions, 17
Composition of limits theorem, 162
confluent hypergeometric limit function, 438
Congruent modulo n, 47
Conjugate in $\mathbb{Z}[\sqrt{d}]$ or $\mathbb{Q}[\sqrt{d}]$, 449
Connected set, 173
Connectedness lemma, 174
Constant function, 20
Constant(s)
  Euler-Mascheroni γ, 153, 193
Continued fraction, 123
  canonical, 415
  unary, 422
Continued fraction convergence theorem, 416
Continued fraction(s)
  finite, 391, 393
  infinite, 393
  nonnegative, 404
  regular, 411
  simple, 392
  terminating, 393
Continued fractions
  transformation rules of, 394
Continuity
  at a point, 166
  on a set, 167
Continuity theorem for power series, 295
Continuous function(s)
  algebra of, 168
  composition of, 168
Continuous functions, 166
Contractive sequence, 120
Contractive sequence theorem, 120
Contrapositive, 6, 11, 24
Convergence
  double sequences, 300
  double series, 302
  functions, 154
  infinite products, 351
  infinite series, 124
  sequences, 95
Convergent of a continued fraction, 393
Converse, 13
Convex set, 72
Coprime numbers, 47, 153, 234, 378
Cosecant function, 200
  power series, 331
Cosine function, 144, 198
Cotangent
  continued fraction, 445
Cotangent function, 200
  power series of, 329
Countability
  of algebraic numbers, 89
  of rational numbers, 86
Countable
  set, 83
Countably infinite, 83
Cover of sets, 172
Cube root, 57

d'Alembert's ratio test, 284
d'Alembert, Jean Le Rond, 284
  quote, 95
Davis' Broadway cafe, 419
de Lagny, Thomas, 343
de Moivre's formula, 198
de Moivre, Abraham, 198
De Morgan and Bertrand's test, 293
De Morgan, Augustus, 293
  laws, 9
  quote, 35
Decimal system, 49
Decimal(s)
  integers, 49
  real numbers, 146
Degree, 202
Dense, 170
Density of the (ir)rationals, 69
Difference of sets, 7
Difference sequence, 122
Digit(s), 49, 147
Diophantine equations, 407
Diophantus of Alexandrea, 389, 407

Dirac, Paul, 237
Direct proof, 24
Dirichlet eta function, 376
Dirichlet function, 17, 159
  modified, 169
Dirichlet's approximation theorem, 430
Dirichlet's test, 273
Dirichlet's theorem for rearrangements, 314
Dirichlet, Johann, 17, 159, 273
Disconnected set, 173
Discontinuity
  jump, 183
Disjoint sets, 7
Distance between vectors or points, 76
Distributive law
  complex numbers, 80
  integers, 36
  natural numbers, 22
  reals, 54
  vectors, 73
Divergence
  infinite products, 351
  infinite series, 124
  proper (for functions), 164
  proper (for sequences), 109
  sequences, 95
  to $\pm\infty$ for functions, 164
  to $\pm\infty$ for sequences, 109
  to zero for infinite products, 351
Divide, divisible, 41
Divisibility rules, 42
Divisibility tricks, 53
Division algorithm, 42
Divisor, 41
Domain of function, 16
dot product, 73
Double a cube, 228
Double sequence, 145, 300
Duodecimal system, 52
Dux, Erich, 321
Dyadic system, 50

e, 93, 114
  approximation, 143
  infinite nested product formula, 143
  irrationality, 143, 276, 420
  nonsimple continued fraction, 389, 402
  simple continued fraction, 389, 419, 443
e and π in a mirror, 386
Eddington, Sir Arthur, 21
Einstein, Albert, 376
Empty set, 4
Enharmonic Harmonium, 437
$\varepsilon/2$-trick, 102, 119
ε-δ arguments, 156
ε-δ definition of limit, 154
ε-N arguments, 96
ε-N definition of limit, 94
ε-principle, 102
Equality
  of functions, 20
  of sets, 5
Eratosthenes of Cyrene, 456
Erdös, Paul, 325
Euclid's theorem, 44
Euclidean algorithm, 46
Euclidean space, 72
Eudoxus of Cnidus, 68
Euler numbers, 270, 330
Euler's identity, 198
Euler's sum for $\pi^2/6$, 126
Euler, Leonhard, 126, 153, 234, 249, 339, 341, 389, 399, 441, 462, 464
  e, 114
  $f(x)$ notation, 16
  identity for e^{iz}, 198
  letter to a princess, 435
  on the series $1 - 1 + 1 \cdots$, 93
  on transcendental numbers, 88
  popularization of π, 228
  quote, 41, 93
  role played in FTA, 210
  summation notation Σ, 28
Euler-Mascheroni constant, 153
Euler-Mascheroni constant γ, 193
Even
  function, 299
  number, 44
Existence
  of complex n-th roots, 215
  of real n-th roots, 65
Exponent, 28, 190, 222
Exponential function, 94, 140, 187, 285
  application of Cauchy multiplication, 319
  the most important function, 153
Exponential funtion
  continued fraction, 442
Extended real numbers, 109, 280

Factor, 41
Factorial(s), 31, 46
  how to factor, 48
Family of sets, 8
Fandreyer, Ernest, 210
Ferguson, D., 234
Fermat's last theorem, 462
Fermat's theorem, 47
Fermat, Pierre, 462
Fermat, Pierre de, 47
Fibonacci sequence, 34, 131, 278, 289, 341, 411
Fibonacci, Leonardo, 341, 418
Field, 54
  ordered, 54
Finite, 83
  subcover, 172

subcover property, 173
Flajolet, Phillip, 339, 341
FOIL law, 24
Formula(s)
$2 = 2 = \frac{e^1}{e^{1/2}} \cdot \frac{e^{1/3}}{e^{1/4}} \cdots$, 194
$\frac{\pi}{4} = \sum_{n=0}^{\infty} \arctan\left(\frac{1}{F_{2n+1}}\right)$, 342
$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^z} = \frac{3^z}{3^z+1} \cdot \frac{5^z}{5^z-1} \cdots$, 386
$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^{2k+1}} = (-1)^k \frac{E_{2k}}{2(2k)!} \left(\frac{\pi}{2}\right)^{2k+1}$, 385
$e = 2 + \frac{2}{2\,+}\,\frac{3}{3\,+}\,\frac{4}{4} \ldots$, 402
addition for (co)sine, 199
$\arctan x = \frac{x}{1\,+}\,\frac{x^2}{3-x^2\,+}\,\frac{3^2x^2}{5-3x^2\,+} \ldots$, 398
Castellano's, 342
change of base, 196
continued fraction for e^x, 442
cosecant power series, 331
cotangent continued fraction, 445
cotangent power series, 329
de Moivre's, 198
double angle for (co)sine, 199
e nonsimple continued fraction, 389
e simple continued fraction, 389, 443
$\frac{e}{e-1} = \lim_{n\to\infty} \left\{ \left(\frac{n}{n}\right)^n + \cdots + \left(\frac{1}{n}\right)^n \right\}$, 191
$e = \lim_{n\to\infty} \left\{ \left(\frac{2}{1}\right)^1 \left(\frac{3}{2}\right)^2 \cdots \left(\frac{n+1}{n}\right)^n \right\}^{1/n}$, 345
$e = \frac{2}{1} \cdot \frac{5}{4} \cdot \frac{16}{15} \cdots$, 352
$\frac{e^{\pi}-e^{-\pi}}{2\pi} = \prod_{n=1}^{\infty} \left(1 + \frac{1}{n^2}\right)$, 362
$\frac{e^{2/x}+1}{e^{2/x}-1} = x + \frac{1}{3x\,+}\,\frac{1}{5x} \ldots$, 441
Euler's $(\pi^2/6)$, 154, 234, 269
Euler's $(\pi^2/6)$, Proof I, 249
Euler's $(\pi^2/6)$, Proof II, 251
Euler's $(\pi^2/6)$, Proof III, 253
Euler's $(\pi^2/6)$, Proof IV, 257
Euler's $(\pi^2/6)$, Proof V, 258
Euler's $(\pi^2/6)$, Proof VI, 264
Euler's $(\pi^2/6)$, Proof VII, 345
Euler's $(\pi^2/6)$, Proof VIII, 370
Euler's $(\pi^2/6)$, Proof VIIII, 371
Euler's $(\pi^2/6)$, Proof X, 372
Euler's $(\pi^2/6)$, Proof VIIIIII, 383
Euler's $(\pi^4/90)$, 262, 371, 383
Euler's $(\pi^6/945)$, 262, 371, 383
Euler's formula $(\eta(2k))$, 383
Euler's formula $(\zeta(2k))$, 263, 383
Euler's infinite product $(\cos \pi z)$, 364
Euler's infinite product $(\cos \pi z)$, Proof I, 364
Euler's infinite product $(\cos \pi z)$, Proofs II,III,IV, 364
Euler's infinite product $(\pi^2/15)$, 384
Euler's infinite product $(\pi^2/6)$, 350, 375, 383
Euler's infinite product $(\pi^4/90)$, 384
Euler's infinite product $(\sin \pi x)$, 238
Euler's infinite product $(\sin \pi z)$, 349, 359
Euler's infinite product $(\sin \pi z)$, Proof I, 241
Euler's infinite product $(\sin \pi z)$, Proof II, 249
Euler's infinite product $(\sin \pi z)$, Proof III, 361
Euler's infinite product $(\sin \pi z)$, Proof IV, 362
Euler's partial fraction $(\frac{\pi}{4\cos\frac{\pi z}{2}})$, 369
Euler's partial fraction $(\frac{\pi}{\sin \pi z})$, 366, 369
Euler's partial fraction $(\pi \tan \frac{\pi z}{2})$, 369
Euler's partial fraction $(\pi z \cot \pi z)$, 366
Euler's partial fraction $(\pi/\sin \pi z)$, 350
Euler's product $(\frac{\pi}{2} = \frac{3}{2} \cdot \frac{5}{6} \cdot \frac{7}{6} \cdots)$, 350, 386
Euler's product $(\frac{\pi}{4} = \frac{3}{4} \cdot \frac{5}{4} \cdot \frac{7}{8} \cdots)$, 350, 386
$\gamma = 1 - \sum_{n=2}^{\infty} \frac{1}{n}\big(\zeta(n) - 1\big)$, 339
$\gamma = \frac{3}{2} - \log 2 - \sum_{n=2}^{\infty} \frac{(-1)^n}{n} (n-1)\big(\zeta(n) - 1\big)$, 339
$\gamma = \sum_{n=2}^{\infty} \frac{(-1)^n}{n} \zeta(n)$, 269, 339
$4/\pi$ continued fraction, 389
Gregory-Leibniz-Madhava, 239
Gregory-Leibniz-Madhava's, 154, 234, 269
Gregory-Leibniz-Madhava, Proof I, 254
Gregory-Leibniz-Madhava, Proof II, 340
Gregory-Leibniz-Madhava, Proof III, 386
Gregory-Madhava's arctangent, 338
half-angle for (co)sine, 200
hyperbolic cotangent continued fraction, 439
hyperbolic secant power series, 330
hyperbolic tangent continued fraction, 441
$\log(1+x) = \frac{x}{1\,+}\,\frac{1^2x}{(2-1x)\,+}\,\frac{2^2x}{(3-2x)} \ldots$, 402
$\log 2$
$\log 2 = \frac{1}{1\,+}\,\frac{1^2}{1\,+}\,\frac{2^2}{1\,+} \ldots$, 396
$\log 2$
$\log 2 = \sum_{n=2}^{\infty} \frac{1}{2^n}\, \zeta(n)$, 309
Lord Brouncker's, 233, 389, 397
Machin's, 234, 270, 342
partial fraction of $1/\sin^2 x$, 257
partial fraction of $1/\sin^2 x$, Proof I, 257
partial fraction of $1/\sin^2 x$, Proof II, 257
$\Phi = \lim_{n\to\infty} \frac{F_{n+1}}{F_n}$, 422
$\Phi = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{F_n F_{n+1}}$, 422
$\Phi^{-1} = \sum_{n=2}^{\infty} \frac{(-1)^n}{F_n F_{n+2}}$, 422
$\pi = 3 + \frac{1^2}{6\,+}\,\frac{3^2}{6\,+}\,\frac{5^2}{6} \ldots$, 400
$\pi \cot \pi x = \frac{1}{x\,+}\,\frac{x^2}{1-2x\,+}\,\frac{(1-x)^2}{2x} \ldots$, 403
$\frac{\pi}{2} = 1 + \frac{1}{1\,+}\,\frac{1\cdot 2}{1\,+}\,\frac{2\cdot 3}{1} \ldots$, 399
$\frac{\cos\frac{\pi x}{2}}{\frac{\pi}{2}} = x + 1 + \frac{(x+1)^2}{-2\cdot 1\,+}\,\frac{(x-1)^2}{-2} \ldots$, 403
$\frac{\sin \pi x}{\pi x} = 1 - \frac{x}{1\,+}\,\frac{(1-x)^2}{2x\,+}\,\frac{(1+x)^2}{1-2x} \ldots$, 403

$\frac{\sin \pi x}{\pi x} = 1 - \frac{x}{1+}\,\frac{1\cdot(1-x)}{x+}\,\frac{1\cdot(1+x)}{1-x}\ldots$, 403
$\frac{\tan \pi x}{\pi x} = 1 + \frac{x}{1-2x+}\,\frac{(1-x)^2}{2x}\ldots$, 403
secant power series, 330
Seidel's, 349, 353
Seidel's for $\frac{\log \theta}{\theta - 1}$, 354
$\frac{6}{\pi^2} = 0^2 + 1^2 - \frac{1^4}{1^2+2^2+}\,\frac{-2^4}{2^2+3^2+}\,\frac{-3^4}{3^2+4^2}\cdots$, 399
$\frac{\pi^2}{6} = \frac{1}{0^2+1^2+}\,\frac{-1^4}{1^2+2^2+}\,\frac{-2^4}{2^2+3^2+}\,\frac{-3^4}{3^2+4^2}\cdots$, 399
$\frac{6}{\pi^2} = 1 - \frac{1}{2^2} - \frac{1}{3^2}\cdots + \frac{\mu(n)}{n^2} + \cdots$, 384
Sondow's, 247
$\sqrt{\pi} = \lim_{n\to\infty} \frac{(n!)^2\, 2^{2n}}{(2n)!\,\sqrt{n}}$, 248
tangent continued fraction, 445
tangent power series, 329
Viète's, 154, 233, 237, 240
Wallis', 234, 237, 247
$\frac{1}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^z}$, 375
$\zeta(z) = \prod \left(1 - \frac{1}{p^z}\right)^{-1}$, Proof I,II, 373
$\zeta(z) = \prod \left(1 - \frac{1}{p^z}\right)^{-1}$, Proof III, 379
$\zeta(z)^2 = \sum_{n=1}^{\infty} \frac{\tau(n)}{n^z}$, 380
$\frac{\zeta(2z)}{\zeta(z)} = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^z}$, 380
$(k+2)\,\zeta(k+1) = \sum_{\ell=1}^{k-2} \zeta(k-\ell)\,\zeta(\ell+1) + 2\sum_{n=1}^{\infty} \frac{H_n}{n^k}$, 265
$\zeta(k) = \sum_{\ell=1}^{k-2} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{m^{\ell}(m+n)^{k-\ell}}$, 310
Fraction rules, 56
Function(s)
absolute value, 17
additive, 171
bijective, 18
characteristic, 20
codomain of, 16
component, 161
composition of, 17
confluent hypergeometric limit function, 438
constant, 20
continuous, 166
cosecant, 200
cosine, 144, 198
cotangent, 200
definition, 16
Dirichlet, 17, 159
Dirichlet eta, 376
domain of, 16
equal, 20
even, 299
exponential, 94, 140, 187, 285
exponential, continued fraction, 442
graph of, 16
greatest integer, 69
Hurwitz zeta, 381
hyperbolic, 208
hyperbolic cosine, 200
hyperbolic sine, 200
hypergeometric, 438
identity, 20
image of, 16
injective, 18
inverse, 18
inverse or arc cosine (complex), 226
inverse or arc cosine (real), 218
inverse or arc sine (complex), 226
inverse or arc sine (real), 218
inverse or arc tangent (complex), 223
inverse or arc tangent (real), 218
inverse or arc tangent continued fraction, 398
jump, 186
Liouville's, 380
logarithm, 189
Möbius, 375, 379
monotone, 182
multiple-valued, 219
multiplicative, 171
nondecreasing, 182
odd, 299
one-to-one, 18
onto, 18
range of, 16
Riemann zeta, 192, 269, 285
Riemann zeta (in terms of Möbius function), 375
Riemann zeta (in terms of primes), 373
secant, 200
sine, 144, 198
strictly decreasing, 182
strictly increasing, 182
surjective, 18
tangent, 200
target of, 16
the most important, 153, 187
value of, 16
Zeno, 182
Fundamental recurrence relations, 405
Fundamental theorem
of algebra, 210
of algebra, proof I, 211
of algebra, proof II, 213
of algebra, proof III, 216
of arithmetic, 45

Galois, Evariste, 193, 452
Game
coin, 34
of Nim, 40, 51
Towers of Hanoi, 33
Game of Nim, 40, 51
Gauss' test, 293
Gauss, Carl, 10, 164
fundamental theorem of algebra, 210

on Borchardt's algorithm, 235
quote, 81, 146, 199
Generalized power rules theorem, 190
Geometric mean, 34
Geometric series, 126
Geometric series theorem, 126
Gilfeather, Frank, 323
Glaisher, James
quote, 340
Golden Ratio, 34
Golden ratio, 94, 113, 392, 414
continued fraction, 94, 123, 392
false rumors, 113, 392
infinite continued square root, 94, 113
the "most irrational number, 427
Good approximation, 423
Goodwin, Edwin, 228
Goto, Hiroyuki, 153
Graph, 16
Greatest common divisor, 46
Greatest integer function, 69
Greatest lower bound, 64
Gregorian calendar, 433
Gregory, James, 234, 338, 340
Gregory-Madhava's arctangent series, 338

Hadamard, Jacques, 287
Halmos, Paul, 3
Harmonic
product, 351
series, 125, 130
Hermite, Charles, 271, 445
almost integer, 206
Heron of Alexandria, 79
Heron's formula, 79
Hidden assumptions, 12
High school
graph of the exponential function, 188
horizontal line test, 185
$i = \sqrt{-1}$, 215
logarithms, 196
long division, 149
plane trigonometry facts, 207
trig identities, 199
zeros of (co)sine, 205
Hilbert, David, 389
Hobbes, Thomas, 4
Hofbauer, Josef, 249, 251, 257
Holy bible, 153, 227, 423
House bill No. 246, 153, 228
Hurwitz zeta function, 381
Hurwitz, Adolf, 381
Huygens, Christian, 433, 437
Hyperbolic
cosine, 200
secant, power series of, 330
sine, 200
Hyperbolic cotangent
continued fraction, 439
Hyperbolic tangent
continued fraction, 441
Hypergeometric function, 438

i, 81
I Kings, 227, 423
Identity
Cassini, 411
Euler's, 198
function, 20
Lagrange, 78
Pythagorean, 199
Identity theorem, 298
If ... then statements, 5, 11
If and only if statements, 13
II Chronicles, 227, 423
Image of a set, 18
Image of function, 16
Imaginary unit, 81
Independent events, 377
Index of a polynomial, 89
Induction, 23
Inductive definitions, 28
Inequalities
rules, 38
strict, 23
Inequality
Arithmetic-geometric mean, 34, 195
Bernoulli, 30
Cauchy-Bunyakovskiĭ-Schwarz, 74
Schwarz, 74
Inequality lemma, 174
Infimum, 64
Infinite
countably, 83
interval, 4
limits, 164
product, 240
series, 124
set, 83
uncountably, 83
Infinite product, 102
Injective, 18
inner product, 73
Inner product space, 74
Integer(s), 35
almost, 206
as a set, 4
Intermediate value property, 176
Intermediate value theorem, 176
Intersection
of sets, 7
family of sets, 9
Interval(s)
end points, 4
chains of, 181
closed, 4

  infinite, 4
  left-half open, 4
  nested, 69
  nontrivial, 170
  of music, 435
  open, 4
  right-half open, 4
Inverse hyperbolic functions, 226
Inverse image of a set, 18
Inverse of a function, 18
Inverse or arc cosine (complex), 226
Inverse or arc cosine (real), 218
Inverse or arc sine (complex), 226
Inverse or arc sine (real), 218
Inverse or arc tangent (complex), 223
Inverse or arc tangent (real), 218
Irrational number, 55
Irrationality
  of e^r for $r \in \mathbb{Q}$, 441
  of e, proof I, 143
  of e, proof II, 276
  of e, proof III, 420
  of $\log r$ for $r \in \mathbb{Q}$, 446
  of logarithmic numbers, 62
  of $\sqrt{2}$, 57, 60, 63
  of trigonometric numbers, 60, 63
Isolated point, 167
Iterated
  limits, 301
  series, 145, 302

Jones, William, 153, 228
Julian calendar, 433
Jump, 183
  discontinuity, 183
  function, 186

Kanada, Yasumasa, 153, 234, 259, 340
Kasner's number, 117
Kasner, Edward, 117
Khayyam, Omar, 434
King-Fang, 437
Knopp, Konrad, 344
Knott, Ron, 417
Kornerup, 437
Kortram, R.A., 372
k-th power free, 381
Kummer's test, 290
Kummer, Ernest, 290

L'Hospital's rule, 195
Lagrange identity, 78
Lagrange, Joseph-Louis, 450
Lambert, Johann, 153, 228, 441
Landau, Edmund, 291
Lange, Jerry, 399
Law of cosines, 78
Law of sines, 79
Leading coefficient, 59
Leap year, 433
Least upper bound, 63
Left-hand limit, 163
Lehmer, D.H., 342
Leibniz, Gottfried, 79, 234, 239, 340
  function word, 15
  on the series $1 - 1 + 1 \cdots$, 93
  quote, 435
Lemma
  Abel's, 271
  compactness, 173
  connectedness, 174
  inequality, 174
Length
  Euclidean, 74
  of complex numbers, 82
Limit
  infimum, 281
  supremum, 280
Limit comparison test, 137
$\liminf$, 281
Limit points and sequences lemma, 155
$\limsup$, 280
Limit(s)
  iterated, 301
  left-hand, 163
  of a function, 155
  of a sequence, 95
  open ball definition for functions, 156
  open ball definition for sequences, 96
  point, 154
  right-hand, 163
Limits
  at infinity, 163
Lindemann, Ferdinand, 153, 228
linear polynomial, 88
Liouville number, 467
Liouville's function, 380
Liouville's theorem, 467
Liouville, Joseph, 380, 467
$\log 2$, 154, 194
  $\log 2 = \frac{1}{1+}\frac{1^2}{1+}\frac{2^2}{1+}\ldots$, 396
  $\log 2 = \sum_{n=2}^{\infty} \frac{1}{2^n}\zeta(n)$, 309
  as alternating harmonic series, 154, 194
  rearrangement, 311
Logarithm
  common, 62
  complex, 221
    power series, 336
  function, 189
  general bases, 196
  natural, 189
Logarithmic test, 294
Logical quantifiers, 14
Lower bound, 64
  greatest, 64
Lucas numbers, 418

Lucas, François, 418

Möbius function, 375
Möbius inversion formula, 379
Möbius, Ferdinand, 375
Machin's formula, 342
Machin, John, 153, 228, 234, 342
Madhava of Sangamagramma, 234, 332, 338, 340
Massaging (an expression), 97, 117, 156, 157
Mathematical induction, 27
Max/min value theorem, 175
Maximum
  of a set, 70, 113
  strict, 180
Mazur, Marcin, 63
Measurement of the circle, 226, 228
Meister, Gary, 323
Mengoli, Pietro, 249
Mercator, Gerhardus, 437
Mercator, Nicolaus, 332
Mersenne number, 48
Mersenne prime, 48
Mersenne, Marin, 48
Mertens' multiplication theorem, 316
Mertens, Franz, 316
Method
  AGM (to compute π), 233
  bisection, 181
  of partial fractions, 125, 128, 210
Minimum
  of a set, 70
MIT cheer, 198
Mnemonic
  π, 153
Modular arithmetic, 47
Monotone
  criterion, 111
  function, 182
  sequence, 111
Monotone criterion theorem, 112
Monotone inverse theorem, 185
Monotone subsequence theorem, 113
"Most"
  extreme irrational, 468
  important equation $e^{i\pi} + 1 = 0$, 203
  important function, 153, 187
  irrational number, 427
  real numbers are transcendental, 83
Multinomial theorem, 34
Multiple, 41
Multiple-valued function, 219
Multiplicative function, 171
Multiplicative identity
  existence for complex numbers, 80
  existence for integers, 36
  existence for natural numbers, 22
  existence for reals, 54
  vectors, 73
Multiplicative inverse
  existence for complex numbers, 80
  existence for reals, 54
Multiplicity of a root, 88
Multiply by conjugate trick, 157
Music, 435

$\mathbb{N}_n$, 83
n-th term test, 124
Natural numbers ($\mathbb{N}$), 22
  as a set, 4
Negation of a statement, 14
Nested intervals, 69
Nested intervals theorem, 70
Newcomb, Simon, 226
Newton, Sir Isaac, 332
Niven, Ivan, 322
Nondecreasing
  function, 182
  sequence, 111
Nonincreasing
  function, 182
  sequence, 111
Nonnegative series test, 125
Nontrivial interval, 170
Norm, 76
  ball, 76
  box, 79
  Euclidean, 74
  sup (or supremum), 79
Normed space, 76
Null sequence, 96
Number theorem series, 310
Number(s)
  algebraic, 88
  approximable, 464
  Bernoulli, 262, 270, 327
  composite, 44
  coprime, 47, 153, 234, 378
  Euler, 270, 330
  extended real, 109, 280
  Liouville, 467
  perfect, 48
  prime, 44
  relatively prime, 47, 234, 378
  square-free, 153, 322
  transcendental, 88

Odd
  number, 44
  function, 299
$1/n$-principle, 67
One-to-one, 18
Onto, 18
Open
  ball, 77
  interval, 4
  set, 173

Order laws
  integers, 38
  natural numbers, 22
  reals, 55
Ordered field, 54
Orthogonal, 77
Oscillation theorem, 203

p-series, 133, 134
p-test, 134
Pandigital, 432
Pappus of Alexandria, 456
Paradox
  Russell, Bertrand, 11
  Vredenduin's, 90
Parallelogram law, 77
Partial
  products, 351
  sum, 124
Partial quotients, 416
Pascal's method, 34
Pascal's rule, 33
Pascal, Blaise, 349
Peirce, Benjamin, 203
Pell equation, 458
Pell, John, 459
Perfect number, 48
Period
  of a b-adic expansion, 149
  of a continued fraction, 447
Periodic
  b-adic expansions, 149
  continued fraction, 447
  purely periodic continued fraction, 452
Persian calendar, 437
Pfaff, Johann, 235
Φ, *see also* Golden ratio
π, 198, 226
  and the unit circle, 204
  continued fraction, 400, 419
  definition of, 202
  formulas, 233
  origin of letter, 228
  Viète's formula, 154, 233, 237
Pianos, 435
Pigeonhole principle, 84
Pochhammer symbol, 438
Pochhammer, Leo August, 438
Poincaré, Henri, 26, 189
Point
  isolated, 167
  limit, 154
Pointwise discontinuous, 170
Polar decomposition, 83
Polar representation, polar coordinates, 206
Polynomial
  complex, 87
  degree of, 59
  index of, 89
  leading coefficient of, 59
  linear, 88
Pope Gregory XIII, 433
Positivity property
  integers, 36
  reals, 54
Power
  complex, 190, 222
  rules (generalized), 190
  rules (integer powers), 57
  rules (rational powers), 67
Power series, 287, 295
  composition of, 325
Power series composition theorem, 325
Power series division theorem, 327
Preimage of a set, 18
Preservation of inequalities theorem
  limits of functions, 162
  limits of sequences, 106
Prime(s)
  definition, 44
  infinite series of, 270, 320
  infinitude, 44
  sparseness, 46
Primitive Pythagorean triple, 47, 432
Principal
  argument, 219
  inverse hyperbolic cosine, 226
  inverse hyperbolic sine, 226
  inverse or arc cosine, 226
  inverse or arc sine, 226
  inverse or arc tangent, 223
  logarithm, 221
  n-th root, 215
  value of a^z, 222
Principle
  ε, 102
  of mathematical induction, 27
  $1/n$, 67
  pigeonhole, 84
Principle of mathematical induction, 27
Pringsheim's theorem
  for double sequences, 302
  for double series, 302
  for sequences, 137
Pringsheim, Alfred, 302
Probability, 377
  number is square-free, 234, 378
  numbers being coprime, 234, 378
Product
  Cartesian, 15
  Cauchy, 315
  of sequences, 108
Proof
  by cases, 39
  contradiction, 25
  contrapositive, 24, 39

direct, 24
Properly divergent
functions, 164
sequences, 109
Property(ies)
intermediate value, 176
of Arctan, 224
of lim inf/sup, 281
of lim inf/sup theorem, 281
of power series, 295
of sine and cosine, 199
of the complex exponential, 141
of the logarithm, 189
of the real exponential, 188
of zero and one theorem, 39
Purely periodic continued fraction, 452
Puzzle
antipodal points on earth, 180
barber, 3
bent wire puzzle, 179
coconut, 409
irrational-irrational, 195
mountain, 178
numbers one more than their cubes, 179
rational-irrational, 192
square root, 57
Which is larger, π^e or e^π?, 206
Pythagorean identity, 199
Pythagorean theorem, 77
Pythagorean triple, 47, 432, 463

Quadratic irrational, 448
Quote
Abel, Niels, 123
Bernoulli, Jacob, 125, 269
Brahmagupta, 459
Cantor, Georg, 83
d'Alembert, Jean Le Rond, 95, 284
Dirac, Paul, 237
Eddington, Sir Arthur, 21
Euler, Leonhard, 41, 93
Galois, Evariste, 193
Gauss, Carl, 10, 81, 146, 164, 199
Glaisher, James, 340
Halmos, Paul, 3
Hermite, Charles, 271
Hilbert, David, 389
Hobbes, Thomas, 4
Leibniz, Gottfried, 79, 435
MIT cheer, 198
Newcomb, Simon, 226
Peirce, Benjamin, 203
Poincaré, Henri, 26, 189
Ramanujan, Srinivasa, 432
Richardson, Lewis, 11
Russell, Bertrand, 11
Schrödinger, Erwin, 55
Smith, David, 153
Wigner, Eugene, 79
Zagier, Don Bernard, 44
Quotient, 41, 42
of sequences, 108
Quotients
complete, 416
partial, 416

Raabe's test, 291
Raabe, Joseph, 291
Radius of convergence, 287
of (co)tangent, 388
Ramanujan, Srinivasa
approximation to π, 432
quote, 432
Range of function, 16
Ratio comparison test, 138
Ratio test
for sequences, 111
for series, 284
Rational numbers
as a set, 4
Rational numbers ($\mathbb{Q}$), 54
Rational zeros theorem, 59
Real numbers ($\mathbb{R}$), 53
Rearrangement, 311
Recurrence relations
fundamental, 405
Wallis-Euler, 404
Reductio ad absurdum, 25
relatively prime numbers, 47, 234, 378
Remainder, 42
Remmert, Reinhold, 210
Rhind (or Ahmes) papyrus, 227
Rhind, Henry, 227
Richardson, Lewis, 11
Riemann zeta function, 192, 269, 285, 373, 375
infinite product, 350
Riemann's rearrangement theorem, 312
Right-hand limit, 163
Ring, 36
Root
of a complex number, 210
of a real number, 57
or zero, 178
principal n-th, 215
rules, 66
Root test
for sequences, 110
for series, 284
Roots of polynomials, 87
Rule
L'Hospital's, 195
Rules
absolute value, 39
divisibility, 42
fraction, 56

  inequality, 38
  of sign, 37
  power (general), 190
  power (integers), 57
  power (rational), 67
  root, 66
Russell's paradox, 11
Russell, Bertrand
  barber puzzle, 3
  paradox, 11
  quote, 11

s-adic expansions, 151
Salamin, Eugene, 233
Schrödinger, Erwin, 55
Schwartz inequality, 77
Schwarz inequality, 74
Schwarz, Hermann, 74
Searcóid, Mícheál, 213
Secant function, 200
  power series, 330
Seidel, Ludwig, 349, 353
Semiperimeter, 79
Septimal system, 50
Sequence, 17
  Cauchy, 117
  contractive, 120
  definition, 94
  difference, 122
  double, 145, 300
  Fibonacci, 34, 131, 289, 341, 411
  Lucas, 418
  monotone, 111
  nondecreasing, 111
  nonincreasing, 111
  null, 96
  of bounded variation, 272
  of sets, 8
  strictly decreasing, 114
  strictly increasing, 114
  subsequence of, 106
  tail, 103
Sequence criterion
  for continuity, 167
  for limits of functions, 159
Series
  absolute convergence, 135
  alternating harmonic, 136
  binomial, 333
  geometric, 126
  Gregory-Leibniz-Madhava, 239, 340
  harmonic, 125, 130
  involving number and sum of divisors, 310
  iterated, 145, 302
  Leibniz's, 239, 340
  power, 287, 295
  telescoping, 128
Set(s)
  compact, 173
  complement of, 8
  connected, 173
  countable, 83
  countably infinite, 83
  cover of, 172
  definition, 4
  dense, 170
  difference of, 7
  disconnected, 173
  disjoint, 7
  empty, 4
  family of, 8
  finite, 83
  image of, 18
  infimum of, 64
  infinite, 83
  intersection of, 7
  inverse image of, 18
  least upper bound of, 63, 64
  limit point of, 154
  lower bound of, 64
  maximum of, 113
  open, 173
  preimage of, 18
  supremum of, 63
  uncountable, 83
  union of, 7
  upper bound of, 63
  Venn diagram of, 8
  zero, 171
Shanks, William, 234
Sharp, Abraham, 343
Sine function, 144, 198
Smith, David, 153
Somayaji, Nilakantha, 340
Sondow, Jonathan, 247
Square root, 57
Square-free numbers, 153, 234, 322, 378
Squaring the circle, 153, 228
Squeeze theorem
  functions, 162
  sequences, 105
Standard deviation, 463
Stirling's formula, weak form, 115
Strict maxima, 180
Strictly decreasing
  sequence, 114
Strictly increasing
  function, 182
  sequence, 114
Subcover, finite, 172
Subsequence, 106
Subset, 5
Sum by curves theorem, 304
$\sum 1/p$, 320
Summation
  arithmetic progression, 29

by curves, 304
by squares, 303
by triangles, 304
definition, 28
geometric progression, 30
of powers of integers, 272, 331
Pascal's method, 34
Summation by parts, 271
Sup (or supremum) norm, 79
Supremum, 63
Surjective, 18

Tøndering, Claus, 433
Tail of a sequence, 103
Tails theorem for sequences, 103
Tails theorem for series, 126
tangent
continued fraction, 445
Tangent function, 200
power series of, 329
Tannery's theorem for products, 359
Tannery's theorem for series, 138
Tannery, Jules, 138
Target of function, 16
Telescoping series, 128
Telescoping series theorem, 128
generalization, 131
Temperment, 436
Terminating
decimal, b-adic expansion, 148
Tertiary system, 51
Test(s)
Abel's, 277
alternating series, 275
Cauchy condensation, 133
comparison, 132
De Morgan and Bertrand's, 293
Dirichlet, 273
Gauss', 293
Kummer's, 290
limit comparison, 137
logarithmic, 294
n-th term test, 124
nonnegative series test, 125
p-test, 134
Raabe's, 291
ratio comparison, 138
ratio for sequences, 111
ratio for series, 284
root for sequences, 110
root for series, 284
The Sand Reckoner, 456
Theorem
Abel's limit, 296
Abel's multiplication, 317
Basic properties of sine and cosine, 199
best approximation, 425, 429
binomial, 31
Bolzano-Weierstrass, 114
boundedness, 175
Cauchy's arithmetic mean, 344
Cauchy's double series, 145, 305
Cauchy's multiplication, 318
Cauchy-Hadamard, 287
Continued fraction convergence, 416
continuity of power series, 295
Dirichlet for rearrangements, 314
Dirichlet's approximation, 430
existence of complex n-th roots, 215
FTA, proof I, 211
FTA, proof II, 213
FTA, proof III, 216
generalized power rules, 190
identity, 298
intermediate value, 176
Liouville's, 467
max/min value, 175
Mertens' multiplication, 316
monotone criterion, 112
monotone inverse, 185
monotone subsequence, 113
oscillation, 203
π and the unit circle, 204
power series composition, 325
power series division, 327
Pringsheim's, for double sequences, 302
Pringsheim's, for double series, 302
Pringsheim's, for sequences, 137
properties of the complex exponential, 141
properties of the logarithm, 189
properties of the real exponential, 188
Riemann's rearrangement, 312
sum by curves, 304
Tannery's for products, 359
Tannery's for series, 138
Thomae's function, 169, 186
Thomae, Johannes, 169
Towers of Hanoi, 33
Transcendental number, 88
Transformation rules, 394
Transitive law
cardinality, 84
for inequalities, 23
for sets, 6
Triangle inequality
for $\mathbb{R}^m$, 75
for integers, 40
for series, 135
Trick(s)
divisibility, 53
$\varepsilon/3$, 277
$\varepsilon/2$, 102, 119
multiply by conjugate, 157
probability that a number is divisible by k, 378
to find N, 97

Trisect an angle, 228
Tropical year, 433
2-series, *see also* Euler's sum for $\pi^2/6$

Unary continued fraction, 422
Uncountability
  of transcendental numbers, 89
  of irrational numbers, 87
  of real numbers, 87
Uncountable, 83
Union
  of sets, 7
  family of sets, 9
Uniqueness
  additive identities and inverses for $\mathbb{Z}$, 37
  multiplicative inverse for $\mathbb{R}$, 56
  of limits for functions, 160
  of limits for sequences, 102
Upper bound, 63
  least, 63

Value of function, 16
Vanden Eynden, Charles, 325
Vardi, Ilan, 339, 341
Vector space, 73
Vectors, 72
Venn diagram, 8
Venn, John, 8
Viète, François, 154, 233, 349
Volterra's theorem, 170
Volterra, Vito, 170
Vredenduin's paradox, 90

Waldo, C.A., 228
Wallis' formulas, 247
Wallis, John, 234, 237, 247
  infinity symbol ∞, 5
Wallis-Euler recurrence relations, 404
Wantzel, Pierre, 228
Weierstrass, Karl, 114
Weinstein, Eric, 432
Weisstein, Eric, 247
Well-ordering (principle) of $\mathbb{N}$, 25
Wigner, Eugene, 79
Wiles, Andrew, 462
Williams' formula, 259
Williams' other formula, 265
Williams, G.T., 258

Yasser, 437

Zagier, Don Bernard, 44
Zeno of Elea, 182
Zeno's function, 182
Zero
  of a function, 178
  set of a function, 171
Zeta function, 134, *see also* Riemann zeta function