Beautiful, Simple, Exact, Crazy

Beautiful, Simple, Exact, Crazy: Mathematics in the Real World

Apoorva Khare and Anna Lachowska

Yale UNIVERSITY PRESS

New Haven and London

Yale University Press books may be purchased in quantity for educational, business, or promotional use. For information, please e-mail sales.press@yale.edu (U.S. office) or sales@yaleup.co.uk (U.K. office).

Printed in the United States of America.

Library of Congress Control Number: 2015935012
ISBN 978-0-300-19089-2 (pbk. : alk. paper)

A catalogue record for this book is available from the British Library.

This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

10 9 8 7 6 5 4 3 2 1

Contents

Preface

The idea for this book arose out of an introductory mathematics course, "Mathematics in the Real World", that the authors co-designed and have taught at Yale and Stanford. The purpose of the course is to familiarize students whose primary interests lie outside of the sciences with the power and beauty of mathematics. In particular, we hope to show how simple mathematical ideas can be applied to answer real-world questions.

Thus we see this as a college-level course book that can be used to teach basic mathematics to students with varying skill levels. We discuss specific and relevant real-life examples: population growth models, logarithmic scales, personal finance, motion with constant speed or constant acceleration, computer security, elements of probability, and statistics. Our goal is to combine the right level of difficulty, pace of exposition, and scope of applications for a curious liberal arts college student to study and enjoy.

Additionally, the book could find use by high school students and by anyone wishing to study independently. The prerequisite is only a high school course in algebra. We hope our book can help readers without extensive mathematical training to analyze datasets and real-world phenomena, and to distinguish statements that are mathematically reasonable from those only pretending to be.

Philosophy and goals

Imagine the following dialog in a high school algebra class.

Teacher: "Find the sum of the geometric series $1 + 5 + 5^2 + \ldots + 5^{20}$."

Student (looking out the window and thinking that life is so much bigger than math, and wondering why the class has to suffer through this tedious, long, and pointless computation): "Can't we just type it into a calculator?"

Whatever the teacher's answer, the student knows that he is right at least on one count – life is indeed bigger than math. It is also bigger than physics, history, and sociology. But for some reason the irrelevance of mathematics is much easier to accept. It is not acceptable to say, "I do not care who was the first president of the United States," but it seems just fine to say, "I do not care about the purpose of a geometric series." It is not acceptable to brag, "I'm illiterate!" But many people feel justified telling friends that they cannot understand mathematics. Yet the consequences of refusing to learn even the most basic mathematical ideas are dire both for the individual and for society.[1] As a society, we need to make collective decisions based on information provided by the media and other sources of variable reliability – and the quality of such decisions depends on our understanding of logic and statistics. As individuals, we have to cope with personal finance and Internet security. We have to know how to estimate the chances an event will occur, and how to interpret lots of new information. From a less practical viewpoint, mathematics adds another dimension (or two, or twenty-three[2]) to the way we see the world, which might be a source of inspiration for a person of any occupation.

The student we just encountered likely will come to college to major in the humanities or social sciences, and is part of the target audience for this book. Thus, our main goal is to convince our reader that mathematics can be easy, its applications are real and widespread, and it can be amusing and inspiring.

To illustrate, let us return to our geometric series example. Mathematics is known to have formulas for everything, so it is not surprising that it has a formula for the sum $1 + 5 + 5^2 + \ldots + 5^n$ for any natural n. What might be more surprising is that this formula requires no derivatives, integrals, or trigonometric identities – and the computation (given in Chapter 9) takes only one line.

Practical applications of this simple mathematical concept are everywhere: for instance, it governs the payment of mortgages and car loans for millions of people. There is more! Everyone knows that $\frac{1}{3} = 0.333333...$, but the fact that *any* repeating decimal, say, 0.765765765..., can be easily written as a fraction of two integers follows from the formula for the geometric series. The same formula lies in the foundation of Zeno's paradox of Achilles and a tortoise, covered in Chapter 10, and two centuries later, another Greek, Archimedes, used a geometric series to calculate the area

[1] We will not elaborate on this here, referring the reader, for example, to *Innumeracy* by J.A. Paulos, published by Hill and Wang, New York, 2001.

[2] According to string theory, the world might have 26 dimensions.

under a parabolic arc. Why should we care? If understanding history is a good enough reason, we can recall that by some accounts, Archimedes also took part in the defense of his town of Syracuse during the siege by the Romans (214-212 BC), and might have used his computation to design a parabolic mirror to set the enemy's fleet on fire. Now, fast-forward to the twentieth century. "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension" was the title of the article, published in *Science* in 1967 by the French-American mathematician Benoit Mandelbrot, that opened a new area of study – fractal geometry. The answer to the question in the title depends on how closely we look, and the closer we look, the greater the coast's length becomes. The same is true for many fractal curves that are easy to describe in words but impossible to draw *precisely*. One of the central tools used to understand these objects is the geometric series.

This is just a glimpse of the scope of applications of one simple mathematical idea – from personal finance, to philosophical puzzles, to fractals, objects of such breathtaking beauty that they make the boundary between art and science disappear.

This example encapsulates our philosophy for the book: we would like to show that there is a lot of simple mathematics relevant to people's everyday lives and their creative aspirations.

There is a way to teach a future artist, and there is a way to teach a layperson to appreciate art. In the latter case, the student is not required to be able to draw and paint like a master, but only to see the beauty in the works of others. Our intention for this book is more ambitious: we hope that it can serve as a guide to the world of mathematics, and also as an inspiration for readers to try their hand at developing and solving mathematical models for their own needs. From there it is only one more step to seeing the world from a mathematical point of view.

The book includes many examples and practice problems designed to gradually build students' proficiency and encourage their involvement with the material. The ultimate goal is to make even the students who are "numerically shy" at the start of a course comfortable applying their mathematical skills in a wide range of situations, from solving puzzles to analyzing statistical data.

We also hope that they become "math-friendly," admitting that mathematics can be interesting and cool. We feel that such comfort levels are indeed achieved when we teach this course at Yale and Stanford – based on our discussions with students, their exams and homework, and their feedback from the course.

Contents and structure

Here are the most important features that distinguish our approach.

First, we neither discuss nor assume any knowledge of calculus or trigonometry. There are quite a few simple concepts in mathematics that do not depend on this particular knowledge but have enormous importance in the real world, as a vast number of pertinent applications show.

Second, our preference is for a self-contained linear exposition, devoid of digressions and asides. We also refrain from presenting an overload of pictures, data analysis, or complicated examples. This book should be viewed as supporting material for a first course in college-level mathematics; there is ample opportunity to analyze more complex examples and phenomena in future studies.

The choice of topics in the book was governed by the following principles:

1. The mathematics involved should be simple, accessible to a student with no experience beyond elementary high school algebra, and explainable within a couple of pages.

2. These simple mathematical concepts should generate a wide range of practical, impressive, or amusing real-world applications.

Accordingly, each chapter of the book, starting with Chapter 2, is divided into two parts:

1. The first, shorter part contains the necessary mathematics: definitions, statements, explanations, examples. This part is supposed to be *studied*, or read slowly, with the reader occasionally doing suggested computations. It can also play the role of lecture notes, if the book is used in teaching a course. Finally, the clearly distinct math section can be used as a reference when reading about applications.

2. The second part, which constitutes most of the chapter, is an exploration of various real-world applications. It contains questions posed and answered by means of the mathematical tools presented in the first part. The second part has little or no mathematical argument and is designed to be read leisurely.

The math sections are there for studying the necessary mathematics. The applications sections are for reading about ways this math can be used. By separating these two activities, we hope to promote learning; instead of students half paying attention while reading fifty pages of a to complete their academic requirements, the book will also significantly benefit them in their future careers.

When the course "Mathematics in the Real World" was introduced at Yale University, it received twice as many applications as there was room for, and the course remains quite popular. In course evaluations and thank-you notes, students mention that the course has helped them acquire basic mathematical literacy and confidence, and instead of being scary, it was an enjoyable experience. We hope that this book will elicit similar reactions from some of our readers.

The course has also been introduced at Stanford University. Our long-term hope is that this book will contribute to the regular curriculum in many colleges and universities.

We also hope that our book will be helpful for those who would like to specialize in natural sciences or economics, but who lack some background in pre-calculus and calculus. For them, the book might serve as a first step on their way to more advanced mathematics.

The book might also be useful for advanced high-school students who are interested in real-world applications of the concepts they learn at school. We estimate that the level of exposition is suitable for most high schools, and the book can be used either by teachers, to supplement the standard program, or by students as extracurricular reading.

In addition, the universal appeal of the topics and the minimal mathematical prerequisites needed to understand this text make for a significantly wider audience. This book should be usable for independent study by busy adults who wish to improve their understanding of math (the book is short!). It might also have some appeal for busy but more experienced math fans, who could leisurely scan it for literary references and unusual applications of math concepts.

Finally, we expect the book to be attractive to international audiences. The authors are of Indian and Russian origin, and we tried to give the book an international flavor. Our examples and cultural references are drawn from all over the world.

In this book we tried to keep a balance between being instructional, being entertaining, and being practical. We hope that this approach will help us promote mathematics as an art, a skill, a language, a way of thinking, a game of puzzles, and in general a worthy activity, to a wide audience of readers.

Acknowledgments

We are grateful to Steve Orszag (1943–2011), who suggested the idea of the course "Mathematics in the Real World" to us, and to Michael Frame at Yale University, who was a source of continuous help and inspiration in developing the course and writing the book. We would like to thank Roger Howe, Andrew Casson, Frank Robinson, and Bill Segraves at Yale, as well as Bala Rajaratnam, Ravi Vakil, and Brian Conrad at Stanford, for their generous assistance in introducing the course at these universities.

We would like to thank our parents and our teachers in high school, college, and graduate school, who taught us to appreciate the depth and beauty of mathematics, and encouraged us, by their own example, to share it with others.

It is a pleasure to thank our students for whom the course was created, whose curiosity and enthusiasm helped shape this book.

We are grateful to our editor, Joseph Calamia, for his careful reading of the manuscript and for multiple suggestions and corrections, which significantly improved the exposition, and to Nancy Hulnick and Mary Pasti, whose thorough editing helped refine the text. We would also like to thank the reviewers of the manuscript for their remarks and suggestions.

Special thanks go out to our families for their continuous support, encouragement, and patience during our work on the book.

Chapter 1

Algebra: The art and craft of computation

In this book, we hope to show you that a wide variety of real-world problems and applications can be tackled systematically, comprehensively, and relatively simply by using just a few mathematical formulas and techniques. In order to introduce the mathematics and then to apply it to the real world, it is essential to be able to work with mathematical expressions and quantities in a systematic manner. Thus, we first need to be comfortable with basic operations like adding or multiplying polynomials; solving equations and systems of (linear or other) equations; and choosing an optimal way to simplify an algebraic expression. Developing these techniques is the goal of this chapter.

Sometimes these techniques produce unexpected results which some of you may have seen as "magic tricks." For instance, you can check that multiplying two consecutive odd numbers (or consecutive even numbers) yields one less than a perfect square (e.g., $5 \cdot 7 + 1 = 36 = 6^2$, $10 \cdot 12 + 1 = 121 = 11^2$). Is this *always* the case, or can we find two consecutive odd or even numbers for which this phenomenon does not occur? Note that it is impossible – even for the biggest computer – to *verify* this for all integers in finite time, because there are infinitely many numbers. But, as we will see, there is a simple way to perform just one calculation – and it will do the job for every single case.

Similarly, multiplying three successive integers and adding the middle integer to this product always yields a perfect cube! Why? Once again, we will see in the exercises in this chapter how one calculation reveals the distributive law again, we get

$$(x^2-2)(x+2)-3x^2+6 = (x^2-2)(x+2)+(x^2-2)\cdot(-3)$$
$$= (x^2-2)(x+2-3) = (x^2-2)(x-1),$$

and so just like long expressions involving numbers (as in Example 1.1), variable expressions can also be simplified.

An advantage of using variables is that once verified, a statement involving variables automatically holds for *every* real value assumed by these variables. For instance, consider the following "magic trick."

Example 1.2. *Get your number back!* Here's a magic trick: start with any number, add 2 to it, multiply the sum by 5, subtract 10 from the product, and divide the difference by 5 – and lo and behold! You get your original number back.

Is it possible to explain what is going on without having to verify whether or not this works for every single starting number?

Solution: Indeed it is. Suppose we start with the number x – as opposed to a specific value, we use a variable because it can be set equal to any number. Then the effect of the given operations on the initial value can be explicitly computed:

$$x \to x+2, \qquad x+2 \to 5(x+2) = 5x+10,$$
$$5x+10 \to (5x+10)-10 = 5x, \qquad 5x \to (5x)/5 = x,$$

and indeed, we get the original number back as the trick claims. (Note that the second operation uses the distributive law.) Thus we have explained the magic trick for every starting number x. Of course, if you want to repeat the calculations corresponding to a specific starting number x, simply set x everywhere above equal to that starting number. □

The above example shows that it is indeed possible to use variables to explain general phenomena, or to solve "word problems" using simple mathematics. We now mention an example which will help us answer the question posed at the beginning of this chapter.

Example 1.3. *The sum times the difference.* One very useful identity involves taking two numbers a, b, and multiplying their sum times their difference. Let us compute the result using the distributive law:

$$(a+b)\cdot(a-b) = a^2 + ba - ab - b^2 = a^2 - b^2.$$

This means that the sum and the difference of *any* two real numbers a and b multiply to yield the difference between their squares.

Conversely, this identity can help compute the product of *any* two numbers: take the square of the number exactly in the middle, and subtract the square of the half of the difference between them. Indeed, if the numbers to multiply are $(a+b)$ and $(a-b)$, then the number in the middle is a, and half of the difference between them is b, and our statement reads

$$(a+b)\cdot(a-b)=a^2-b^2,$$

which is the sum times the difference formula. It works especially well for two integers that are equidistant from a number whose square is easy to compute.

For instance, this identity enables us to compute $297 \cdot 303$ without too much fuss (or a calculator):

$$297\cdot 303=(300-3)\cdot(300+3)=300^2-3^2=90,000-9=89,991.$$

□

Example 1.4. *Multiplying consecutive odd or even numbers.* We now return to the example mentioned at the beginning of the chapter: multiplying two consecutive odd (or even) numbers always seems to yield one less than a perfect square. To check this for every pair of consecutive odd/even numbers at once, we replace these numbers by variables: denote the number between the two consecutive odd/even numbers as n. Then the two numbers that we are to multiply, are $(n+1)$ and $(n-1)$. Their product, by the previous example, is

$$(n+1)\cdot(n-1)=n^2-1^2=n^2-1.$$

This is precisely one less than the square of the middle number n, as we claimed. □

This general calculation also provides a method to compute such products: take the square of the number in between, and subtract 1. For instance, if we want to compute the product of the two successive odd numbers $101\cdot 99$, then we simply use $n=100$ in the above calculation, to get $100^2-1^2=9,999$.

Solving equations

Often in the real world, we see that the same quantity can be expressed in different ways using different physical units. For instance, the same distance can be given in yards, miles, or meters. Or, the same temperature can be expressed in Celsius and in Fahrenheit (or even in Kelvin) degrees. In solving questions involving such changes, a simple tool that is useful is a *linear equation.* Here is an example.

Example 1.5. What temperature has the same numerical value when measured in Fahrenheit and Celsius?

Solution: Denote the unknown temperature in Celsius, say, by T. Then this same temperature in Fahrenheit is given by: $32+\frac{9}{5}T$. The conditions of the problem imply that we are to equate these two expressions. In formulas, we have

$$T = 32 + \frac{9}{5}T \quad \Longrightarrow \quad -32 = \frac{9}{5}T - T = \left(\frac{9}{5} - 1\right)T = \frac{4}{5}T$$
$$\Longrightarrow \; T = (-32)\cdot\frac{5}{4} = -40.$$

We conclude that $-40°\mathrm{F} = -40°\mathrm{C}$ (and no other temperature has this property). □

The above equation $T = 32+\frac{9}{5}T$ is an example of a linear equation. It models a linear relationship, or dependency, between two varying quantities (in this case, the temperature in Celsius and in Fahrenheit). A linear equation is characterized by the way the variable appears in it: it can either stand alone, or be multiplied by a numeric coefficient. The variable can be denoted by any letter, say, x, T, w, y, A, and so on. All other terms of the equation are numbers.

Such dependencies are ubiquitous in the real world – for instance, one can use linear equations to model production costs (involving an overhead amount and manufacturing cost per unit), simple interest in banking, distance traveled by a vehicle or a jogger with constant velocity, or conversions between different units of temperature, money, weight, distance, and so on. For example, one can determine which temperature in Fahrenheit is twice (and which is half) the number that it equals in Celsius. Here is another real-world example which can be solved using linear equations.

Example 1.6. *Taxicab fares.* Suppose in a given city, cab drivers charge an initial fare of \$3, followed by an additional charge of \$2 per mile. What is the fare if the passenger travels for 10 miles? How many miles has the passenger traveled if the fare is \$11? *Answer:* \$23 and 4 miles, respectively. The linear equation here is: $fare = 3 + 2\cdot miles$. □

Another class of equations that is easy to solve involves equating a product to zero. For instance: *find all possible values of the variables* a, b, c *such that* $a \cdot b \cdot c = 0$. The answer is that if three numbers are nonzero, their product cannot be zero; hence at least one of the variables a, b, c must equal zero in this case. Thus, the solution is that $a = 0$ and b, c are arbitrary real numbers; or $b = 0$ and a, c are arbitrary; or $c = 0$ and a, b are arbitrary.

Example 1.7. Using the above idea, find all solutions to the following equations.

1. $(x-1)(x+2) = 0$.
2. $(x-1)^2(x+3)(y-2) = 0$.
3. $y^2 - 1 = 99$.

Solution: The first two equations are easy to solve, using the previous reasoning. Thus for the first equation, either $x - 1 = 0$ or $x + 2 = 0$, whence $x = 1$, or $x = -2$. For the second equation, we write the equation as $(x-1) \cdot (x-1) \cdot (x+3) \cdot (y-2) = 0$. Hence at least one of the four factors is zero, which leads to: $x = 1$, and y is any number, or $x = -3$ and y is any number, or $y = 2$ and x is any number. (We allow for both $x = 1$ and $y = 2$ to occur simultaneously.)

Finally for the last equation, we add 1 to both sides to get: $y^2 = 100$. Hence $y = 10$ or -10. The other way to see this is to use Example 1.3. To do so, subtract 99 from both sides to get:

$$y^2 - 100 = 0 \quad \Longrightarrow \quad y^2 - 10^2 = 0 \quad \Longrightarrow (y+10)(y-10) = 0.$$

Hence we obtain the complete set of solutions: $y = -10$ or $y = 10$. For convenience we write $y = \pm 10$. □

Polynomials: Solving quadratic equations

Now we move from simple dependencies to more involved ones. Using variables allows us to define a useful class of expressions called *polynomials*. A polynomial is an expression in which several (finitely many) powers of x can appear, either alone or multiplied by numbers. For instance, $f(x) = 3x^2 - 1$ is equal to $f(x) = 3 \cdot x^2 + (-1) \cdot x^0$. The number next to each power of x (or of the one variable that is used) is called its *coefficient*. For instance, the coefficients of x^0, x^1, x^2, and x^3 in $3x^2 - 1$ are $-1, 0, 3, 0$, respectively.

The highest power of x whose coefficient in a polynomial $p(x)$ is nonzero is called the *degree* of the polynomial, and is denoted by $\deg(p)$. A polynomial is said to be constant, linear, quadratic, cubic, quartic, and so on, if its degree is, respectively, $0, 1, 2, 3, 4$, and so on. A linear equation can be expressed as the condition that a linear polynomial in the variable equals zero: $ax + b = 0$ for some real numbers a, b. In general, given a polynomial $p(x)$, one is often interested in determining the set of x such that $p(x)$ equals a given value – in other words, solving a polynomial equation.

In the remainder of this chapter, we will learn how to solve some polynomial equations which are more involved than linear equations. For instance, using the distributive law (or the FOIL method), you can check that the first of the equations in Example 1.7 was an example of a second-degree polynomial equation in x – i.e., a *quadratic equation*: $x^2 + x - 2 = 0$. Other examples of quadratic equations are $x^2 - 3x + 2 = 0$, or $2x^2 - 8x + 8 = 0$, or more generally,

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

for some real numbers a, b, c with $a \neq 0$. (Note that if $a = 0$ then the equation becomes a linear equation.)

All quadratic equations have two, one, or no real roots. Here is the formula for the *solutions* to the general quadratic equation $ax^2 + bx + c = 0$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

This rule provides the answer in all cases whenever a is nonzero. Recall that the square root can be taken only of nonnegative values. From the formula, we see that there are three possible cases, depending on the quantity under the square root $b^2 - 4ac$, which is called the *discriminant*.

- If the discriminant $b^2 - 4ac$ is positive, the quadratic equation has two distinct real roots, mentioned in the above formula.
- If the discriminant is zero, i.e., $b^2 = 4ac$, then the quadratic equation has exactly one real (repeated) root: $x = -b/2a$.
- If the discriminant is negative (i.e., $b^2 < 4ac$), there are *no* real roots of the equation.

Let us solve some quadratic equations to illustrate the above formula.

Example 1.8. Find all (real) roots of the following equations:

Quadratic-type equations

In the previous section, we saw how to solve a quadratic equation in a variable x. We can apply the same technique to solve more complicated equations that look similar to quadratic equations, by reducing them to that form. For example, the equation $x^4 - 2x^2 + 1 = 0$ is not quadratic, but *quartic*, i.e., of degree 4. Or $x^{200} - 2x^{100} + 1 = 0$ is in fact an equation of degree 200. However, they both look somewhat similar – and in fact, they can both be solved using the formula for roots of quadratic equations.

Example 1.11. Solve the two equations in the previous paragraph.

Solution: Let us begin with the first equation, $x^4 - 2x^2 + 1 = 0$. Note that x^4 is the square of x^2, so if we denote x^2 by a new variable y, then the equation changes to: $y^2 - 2y + 1 = 0$. Now apply the general formula for the roots of a quadratic equation (or see Example 1.8(2)) to obtain: $y = 1$. Substituting back for x, we obtain that $x^2 = 1$, and finally $x = \pm 1$ are the roots of the equation.

For the second equation, $x^{200} - 2x^{100} + 1 = 0$, we again make a substitution: $y = x^{100}$. This leads us to the same *quadratic* equation in y, as in the preceding paragraph. Therefore $y = 1$, i.e., $x^{100} = 1$. The only real numbers whose 100th power is 1 are: $x = \pm 1$. (Thus there are only two real roots to this equation of degree 200.) □

Here are some other equations that can be solved using the formula for the roots of quadratic equations.

Practice problem 1.12. Solve the following equations.

1. $y^5 - 4y^4 + 3y^3 = 0$. *Solution:* First take y^3 common, so either it is zero or the remaining quadratic factor is zero. Solve it to get: $y = 1, 3$, or $y = 0$.

2. $\frac{1}{y^2} - \frac{1}{y} - 2 = 0$. *Solution:* Set $x = 1/y$; then $x^2 - x - 2 = 0$. Solving, $x = -1, 2$, whence $y = 1/x = -1, 1/2$.

Number of real roots

Looking at the previous two examples, you might ask the following question: how many real nth roots does a real number have? For instance, 8 has two square roots, while -8 has no square roots. On the other hand, both 8 and -8 have exactly one cube root each: 2 and -2, respectively.

Here is the answer: if n is even (square root, fourth root, and so on), a negative number has *no* real nth roots, while a positive number has exactly two real nth roots. (And 0 has exactly one nth root: itself.) On the other hand, if n is odd, then every real number has a unique real nth root.

Solving a quadratic equation by completing the square*

We saw above that, given a quadratic equation $ax^2 + bx + c = 0$ (with $a \neq 0$), the two numbers

$$x_\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

indeed satisfy the equation whenever $b^2 \geq 4ac$. Now we will show that $x_\pm$ are the *only* two possible roots to the equation. Let us first demonstrate how to do this in a special case.

Example 1.13. Solve the equation $x^2 + \frac{5}{2}x - 6 = 0$ without using the general formula.

Solution: Recall that if we can rewrite the equation as a product of factors, we will be able to find the roots by setting each of the factors equal to zero (see Example 1.7). To factorize a general quadratic polynomial, there is a useful trick called *completing the square*, which can be used to write the quadratic polynomial as the sum of a real number and the square of a linear polynomial. The trick relies on the following general formula (which can be verified by using the distributive law): for any two real numbers x, y,

$$(x + y)^2 = x^2 + 2xy + y^2.$$

Now suppose we are given a quadratic polynomial like $x^2 + \frac{5}{2}x - 6$. We first consider only the linear and quadratic terms, namely $x^2 + \frac{5}{2}x$, and ask: what number should one add to this in order to obtain a square of the form $(x+y)^2$? We see that the term $\frac{5}{2}x$ should correspond to the term $2yx$, which means that $y = \frac{5}{4}$. In other words, y is half of the coefficient for x, and we add y^2 to obtain $(x+y)^2$. Thus if we add $(\frac{5}{4})^2 = \frac{25}{16}$, we will get $x^2 + \frac{5}{2}x + \frac{25}{16}$, and you should verify, using the distributive law (i.e., the FOIL method), that this equals $(x + \frac{5}{4})^2$.

Now we obtained a square of a linear polynomial, but we added $\frac{25}{16}$ to our original equation and dropped the last term -6. To recover the original equation, we must compensate for this. Thus, we obtain:

$$0 = x^2 + \frac{5}{2}x - 6 = x^2 + \frac{5}{2}x + \frac{25}{16} \quad - \frac{25}{16} - 6 = \left(x + \frac{5}{4}\right)^2 - \frac{121}{16}.$$

In the last line we used the equality $-6 - \frac{25}{16} = \frac{-96-25}{16} = -\frac{121}{16}$. We notice that $\frac{121}{16}$ is a complete square: $\frac{121}{16} = (\frac{11}{4})^2$.

The second step is to use the formula $a^2 - b^2 = (a+b)(a-b)$ derived in Example 1.3, to obtain:

$$\begin{aligned} 0 &= \left(x + \frac{5}{4}\right)^2 - \frac{121}{16} = \left(x + \frac{5}{4}\right)^2 - \left(\frac{11}{4}\right)^2 \\ &= \left(x + \frac{5}{4} + \frac{11}{4}\right) \cdot \left(x + \frac{5}{4} - \frac{11}{4}\right) = (x+4)\left(x - \frac{3}{2}\right). \end{aligned}$$

We conclude that the only possible solutions are $x = -4$ and $x = \frac{3}{2}$. You can check that the general formula for the roots of a quadratic equation gives the same two solutions. □

The above technique works when the coefficient of x^2 is 1. What about a general quadratic equation $ax^2 + bx + c$ with $a \neq 0$? Simply divide by a first:

Example 1.14. Solve the general quadratic equation $ax^2 + bx + c = 0$, by completing the square.

Solution: First divide both sides by $a \neq 0$ to obtain:

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

As discussed in the previous example, we have to add and subtract the square of the half of the coefficient in front of x, which equals $\left(\frac{b}{2a}\right)^2$. We obtain:

$$0 = x^2 + \frac{b}{a}x + \frac{c}{a} = x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a} = \left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2} + \frac{c}{a}.$$

(Note here that $(b/(2a))^2 = b^2/(2a)^2 = b^2/(4a^2)$.) Now we can combine the last two terms by taking the common denominator:

$$-\frac{b^2}{4a^2} + \frac{c}{a} = \frac{-b^2 + 4ac}{4a^2} = \frac{-(b^2 - 4ac)}{4a^2} = -\left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2.$$

This yields:

$$0 = x^2 + \frac{b}{a}x + \frac{c}{a} = \left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2.$$

Note that this is a difference of squares if and only if the discriminant $(b^2 - 4ac)$ is a nonnegative number, so that we can take the square root. If the discriminant is zero, we get a complete square equal to zero, and a single solution $x = -\frac{b}{2a}$. If it is negative, we have a complete square equal to a negative number, which means that no real x satisfies the equation. Finally, in case the discriminant is positive, we can apply the difference of squares formula. Then we get

$$\begin{aligned} 0 &= \left(x + \frac{b}{2a} - \frac{\sqrt{b^2 - 4ac}}{2a}\right) \cdot \left(x + \frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a}\right) \\ &= \left(x + \frac{b - \sqrt{b^2 - 4ac}}{2a}\right) \cdot \left(x + \frac{b + \sqrt{b^2 - 4ac}}{2a}\right), \end{aligned}$$

and we conclude that the two roots to the quadratic equation are as claimed: $x_{\pm} = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. □

EXERCISES

Question 1.1. Expand or contract (factorize) the following expressions, using the distributive law. All variables below denote real numbers.

1. $(a - 2)(b - 2)$.
2. $(x + 1)(2x - 3)$.
3. $(y^2 + 1)(y^2 - 1)$.

Question 1.2. Expand or contract the following expressions using the distributive law.

1. $(1 + z + z^2 + z^3)(1 - z)$.
2. $T^2 - sT + 2T - 2s$, where s is some fixed real number.
3. $AB - BC + CD - DA$.

Question 1.3. Suppose you consider any two consecutive integers. Then the difference between their squares is always an odd number. Can you prove this fact in general?

Question 1.4. Suppose you multiply any three consecutive integers, and add the middle integer to this product. Then you will always get a perfect cube. Why is this? Denote the middle integer by n and carry out this computation using n to show why this holds for *every* integer n, all at once. *Hint:* Decide for yourself in which order you would like to multiply the three integers.

Question 1.5. Solve the following equations for the unknown variables.

1. $x^4 = 25$.

2. $(x-1)^2(x-2)(x+3) = 0$.

3. $y^2 + 8y + 15 = 0$.

4. $A^2 - 3A - 4 = 0$.

Question 1.6. Solve the following equations for the unknown variables.

1. $y^3 = 27$.

2. $(z-22)^3(2z+27)(z+3)^{17} = 0$.

3. $s^2 + 10s + 20 = 0$.

4. $c^4 + 3c^2 - 4 = 0$. *Hint:* Remember that squares are never negative.

Question 1.7. The quadratic equation $y^2 = 6y$ can be solved directly using the distributive law, and also using the general formula. Solve the equation in both possible ways, and verify for yourself that the set of solutions is the same regardless of how you solve the equation.

Question 1.8. For what value(s) of b does the equation $x^2 = bx - 7$ have a unique solution?

Question 1.9. Solve the quadratic equations.

1. $2\beta^2 + 22\beta + 60 = 0$.

2. $y^2 - 3y + 3 = 0$.

Question 1.10. Solve the quadratic equations.

1. $2x^2 - 4\sqrt{5}x + 10 = 0$.

2. $2z^2 - \pi z + 1 = 0$.

Chapter 2

Velocity: On the road

The simplest kind of motion is the motion of an object along a straight line with a constant speed: a car moving along a road at sixty miles per hour, a pedestrian walking at three miles per hour. However, if multiple objects are moving with constant speeds along the same path, and they are allowed to change direction, the mathematical situation becomes more intricate. In this chapter, we consider a range of examples of systems of moving objects drawn from everyday experience, our imagination, and, in one case, literary fiction.

MATH

The mathematical model of the rectilinear constant speed motion of one object can be formulated in one line:

$$s = v \cdot t,$$

where s is the distance from the starting point covered by the object moving along a straight line with speed v during the time t. The three quantities should be measured in compatible units, meaning that if the distance is in miles (mi), and the time in hours (h), then the speed should be taken in miles per hour (mph). The same formula can be read in two other ways:

$$v = \frac{s}{t} \qquad \text{and} \qquad t = \frac{s}{v}.$$

In fact, this model describes a larger class of processes. If the speed of an object is not constant, but the *average* speed is known to be v, then the distance covered in time t is given by the same formula $s = v \cdot t$. Moreover, the path does not need to be straight. If v is the (average) speed of motion along any given path, then $t = \frac{s}{v}$ is the time needed to cover the distance s measured *along* the path.[1]

Example 2.1. How long does it take to get to a town 230 miles away along a road, driving at an average speed of 60 miles per hour? We compute:

$$t = \frac{s}{v} = \frac{230}{60} \simeq 3.833\,\text{h} = 3\,\text{h}\ 50\,\text{min}.$$

□

Now suppose there are two objects moving with the constant speeds v_1 and v_2 along the same path. Will they ever meet, and how soon?

To answer this question, it is convenient to introduce the notion of *relative speed*, defined as the rate of change of the distance between the two objects. Thus, if the two objects are moving along the same path in the same direction, then the relative speed v_{12} is the difference between their speeds:

$$v_{12} = |v_1 - v_2| = \begin{cases} v_1 - v_2, & \text{if } v_1 \geq v_2; \\ v_2 - v_1 & \text{if } v_1 < v_2. \end{cases}$$

If the faster moving object is behind, they are getting closer at the rate v_{12}, and if the slower one is behind, they are getting farther apart at the same rate.

Figure 2.1: Moving in the same direction.

Now, if the two objects are moving in opposite directions, then the relative speed v_{12} is the sum of their speeds:

$$v_{12} = v_2 + v_1.$$

[1] Motion along a curve requires an acceleration directed to the center of the curve. But we are interested in the speed of motion along a path, and we assume it to be constant.

Depending on the position of the objects, they are getting closer or farther apart at the rate v_{12}.

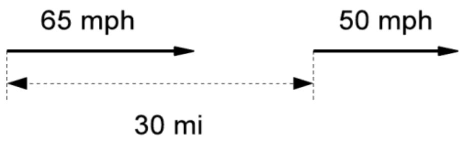
Figure 2.2: Moving in opposite directions.

The time needed for the distance between the two objects to decrease or increase by s is given by the formula

$$t = \frac{s}{v_{12}}.$$

In particular, if the objects are getting closer, and the initial distance between them is s, then $t = \frac{s}{v_{12}}$ is the time before they meet.

Example 2.2. Suppose your friend is 30 miles ahead of you on a highway, moving forward at a constant speed of 50 miles per hour. Can you catch up with him before he gets to the next town, 120 miles ahead of you, without breaking the speed limit of 65 miles per hour? Equivalently, assume you are moving at a constant speed of 65 miles per hour. When will you catch up with your friend: before or after he passes the town 120 miles ahead?



Solution: We have two objects moving in the same direction, with $v_1 = 65$ mph and $v_2 = 50$ mph, the faster object behind. The initial distance between them is $s = 30$ mi. The time before they meet is

$$t = \frac{s}{v_{12}} = \frac{s}{v_1 - v_2} = \frac{30}{65 - 50} = 2 \text{ h}.$$

In 2 hours, you will be $2 \cdot 65 = 130$ miles ahead, already past the town situated 120 miles ahead. You cannot catch up with your friend before he passes the town. □

APPLICATIONS

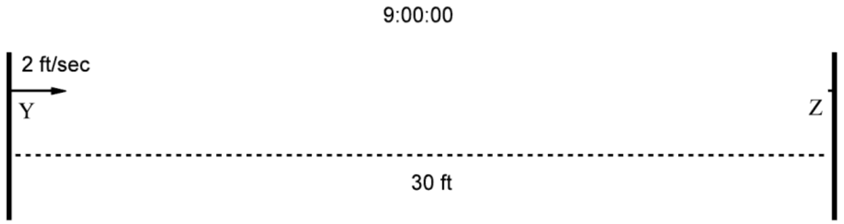
Constant speed motion with a change of direction

If the objects moving along the same path are allowed to change directions, the model becomes more complicated and interesting patterns can emerge. Here are a couple of examples.

Example 2.3. Two bees, Yolanda (Y) and Zoe (Z), fly non-stop between two parallel walls 30 feet (ft) apart. Yolanda starts at 9:00:00 from the western wall and flies with a constant speed of 2 feet per second (ft/sec). Zoe starts at 9:00:10 from the eastern wall and flies with a constant speed of 3 feet per second. When will Yolanda meet Zoe for the first, second, and third time?

Solution: Let us find the time of the first meeting. At 9:00:00, only Yolanda is moving:

Figure 2.3

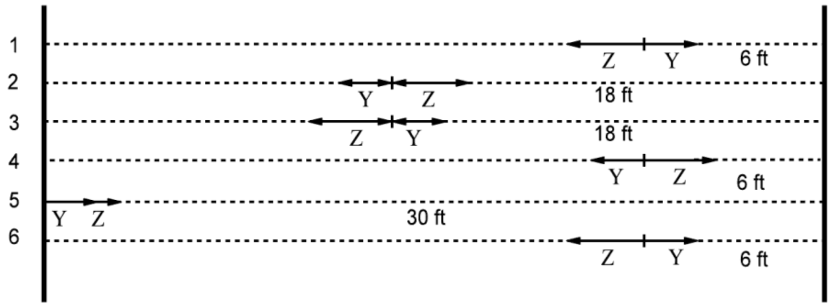


At 9:00:10, when Yolanda is $2 \cdot 10 = 20$ feet from the western wall, Zoe starts flying west at a speed of 3 feet per second from the eastern wall.

To meet, they have to cover the distance $s = 30 - 20 = 10$ ft, moving with the relative speed $v_{12} = v_1 + v_2 = 2 + 3 = 5$ ft/sec. This will take $t = \frac{s}{v_{12}} = \frac{10}{5} = 2$ sec. Therefore, the first meeting will happen 2 seconds after Zoe starts flying, or at 9:00:12.

Clearly, Yolanda and Zoe will meet again if they continue flying between the walls. To figure out the time of the next meeting, we have to (1) find the moment of time when Zoe is at the western wall; (2) find the

Figure 2.6: First six meetings



so on, as long as the bees keep flying. The meetings will take place at 6, 18, or 30 feet from the eastern wall. □

Two trains and a bee

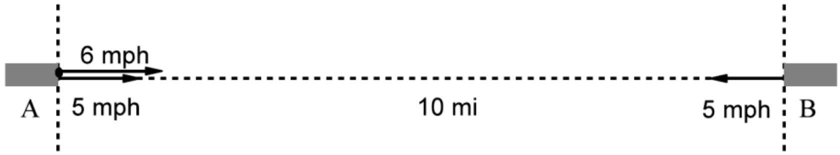
Now imagine that in the previous example the walls were allowed to move as well. A well-known model with just one bee flying between two moving obstacles is considered below.

Example 2.5. Two trains, A from the west and B from the east, are approaching along the same straight train track, each moving at a constant speed of 5 miles per hour. When the distance between the trains is 10 miles, a bee starts flying east from the head of train A. The bee moves at a constant speed of 6 miles per hour until it hits train B, then it turns around and flies west with the same constant speed until it hits train A, and so on until the two trains meet and the bee gets smashed between them (now we are in trouble with the animal rights activists). What is the total distance covered by the bee until the crash?

Solution: We will start with a less ambitious question, namely, when and where will the bee meet train B for the first time? The diagram above shows the initial positions and speeds of the moving objects: the bee goes east at $v_1 = 6$ mph, train B goes west at $v_2 = 5$ mph, and the initial distance between them is $s_0 = 10$ mi. They will meet in

$$t_1 = \frac{10\,\text{mi}}{5 + 6\,\text{mph}} = \frac{10}{11}\ \text{h}.$$

Figure 2.7

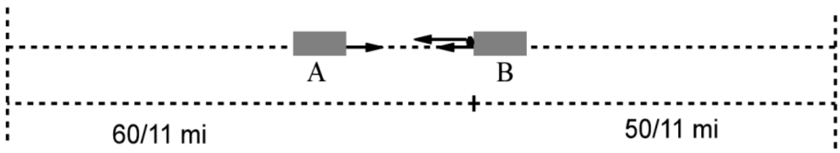


In this time, the bee will cover the distance

$$s_1 = 6\,\text{mph} \cdot \frac{10}{11}\,\text{h} = \frac{60}{11}\,\text{mi}.$$

Now we know the position of the bee at the moment when it changes direction and starts flying west. The diagram shows this moment.

Figure 2.8



Next we can ask when and where the next meeting of the bee with train A will occur. In $\frac{10}{11}$ of an hour, train A has moved $5 \cdot \frac{10}{11} = \frac{50}{11}$ mi east, and is situated $\frac{60}{11} - \frac{50}{11} = \frac{10}{11}$ mi west from the bee. The time to their meeting is

$$t_2 = \frac{\frac{10}{11}\,\text{mi}}{5 + 6\,\text{mph}} = \frac{10}{11 \cdot 11} = \frac{10}{121}\,\text{h}.$$

In this time the bee will cover an additional distance

$$s_2 = 6\,\text{mph} \cdot \frac{10}{121}\,\text{h} = \frac{60}{121}\,\text{mi}.$$

In the first two moves, the bee covered $s_1 + s_2 = \frac{60}{11} + \frac{60}{121} = \frac{660+60}{121} = \frac{720}{121}$ mi. To find the total distance covered by the bee until is gets smashed we would have to compute and sum up infinitely many more such distances, and the whole approach is starting to look somewhat hairy. For the moment we will drop the question, and return to it in Chapter 10, where we will learn how to compute such infinite sums.

However, there is another, more efficient solution: the bee is flying non-stop at a constant speed of 6 miles per hour until the trains meet. The time before the meeting of the trains is

$$t = \frac{10}{5+5} = 1 \text{ h}.$$

Therefore, the total distance covered by the bee before the crash is

$$6\,\text{mph} \cdot 1\,\text{h} = 6\text{ mi}.$$

□

A traveler and messengers

We know of at least one excellent example of an elaborate system of objects moving at constant speed that appears in fiction. Here is a short synopsis of the novella "The Seven Messengers" by an Italian writer, Dino Buzzati.

The prince, the narrator of the story, sets out to explore his father's kingdom, hoping to find its boundaries. He and his knights start from the capital and move in one direction (due south, or so they hope) at a constant speed of 40 leagues per day. After two days of travel, the prince sends his first messenger – Alessandro – back to the capital. The next six messengers, Bartolomeo, Caio, Domenico, Ettore, Federico, and Gregorio, are sent back, respectively, after three, four, five, six, seven, and eight days of travel. All messengers move at a constant speed of 60 leagues per day. Having reached the capital, each messenger immediately starts back along the same path to catch up with the prince. Having reached the prince, the messenger immediately starts back for the capital, and so on. Thus, the seven messengers oscillate between the capital and the prince, while the prince is moving farther and farther away. The intervals between the arrivals of the messengers grow, until one day the prince realizes that the next return of a messenger will be the last one he will live to see. To simplify the diagrams, we assume in Figures 2.9 and 2.10 that the southern direction is to the right.

In the course of the story, the narrator makes multiple numerical statements, and a natural reaction for an inquisitive mind is to check whether or not they make sense. Any mathematician would definitely have an itch to do so; but even a reader with the sole interest in literature, presumably, would like to know if the author is careful about all the details of his creation, or neglectful, or intentionally misleading. On multiple occasions, the narrator mentions the number of days, months, and years that have passed since the beginning of the journey and between the consecutive returns of the messengers.

Figure 2.9: Messenger leaves for the capital.

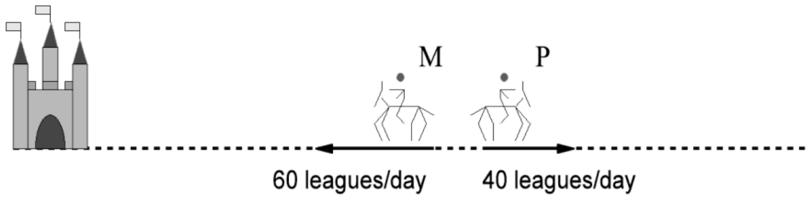
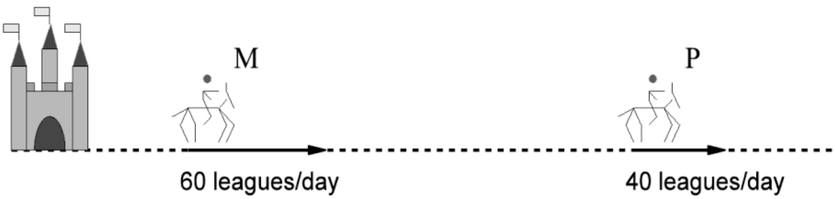


Figure 2.10: Messenger starts back to catch up with the prince.



The interesting feature of this particular work of fiction is that all the numerical statements in it are verifiable. We can figure out the position of each of the messengers at any moment of time (in days since the start of the expedition) and, in particular, check the numerical statements contained in the story.

Let d be the number of days elapsed before a messenger (M) was first sent back. Let t denote the number of days M needs to catch up again with the prince (P). Then we have the equation:

$$\text{distance traveled by } \mathrm{M} = 2\begin{pmatrix}\text{distance to the capital}\\ \text{when M leaves}\end{pmatrix} + $$

$$+\begin{pmatrix}\text{distance traveled by P}\\ \text{since M left}\end{pmatrix}$$

or, taking into account the given speeds of P and M,

$$60t = 2 \cdot 40d + 40t.$$

From this equation we find that $t = 4d$ is the time taken by M to catch up with P. By the time he catches up, $d + 4d = 5d$ days would have passed since the start of the expedition. Now we only need to plug in $d = 2$ for Alessandro, $d = 3$ for Bartolomeo, and so on, up to $d = 8$ for the last messenger, Gregorio, to obtain the times of their return to the prince. We find that they reunite with the prince after 10, 15, 20, 25, 30, 35, and 40 days, respectively.

A messenger is then immediately sent back to the capital; so, for instance, Alessandro is sent back the second time after 10 days since the start of the expedition. By the same argument as before, we see that the second return of the messengers will occur after $5 \cdot 5d = 25d$ since the start of the expedition. Next, they will return after $5 \cdot 25d = 125d$, $5 \cdot 125d = 625d$, and so on. For example, Alessandro ($d = 2$) will return to the camp after $2 \cdot 5 = 10$, $2 \cdot 25 = 50$, $2 \cdot 125 = 250$, $2 \cdot 625 = 1,250$, etc., days.

Here is the timetable (in days since the beginning of the journey) of the first five consecutive returns of each messenger to the prince:

	d	$5d$	$25d$	$125d$	$625d$
Alessandro	2	10	50	250	1,250
Bartolomeo	3	15	75	375	1,875
Caio	4	20	100	500	2,500
Domenico	5	25	125	625	3,125
Ettore	6	30	150	750	3,750
Federico	7	35	175	875	4,375
Gregorio	8	40	200	1,000	5,000

This table contains enough information to check the numerical claims of the narrator. Let us consider each one of them:

1. *"...it was sufficient to multiply by five the days elapsed so far to know when the messenger would catch up with us."* This is exactly the formula we derived above: if a messenger leaves the prince after d days since the start of the expedition, he returns to the prince after $5d$ days.

2. After fifty days, the interval between the messenger's consecutive returns increases to twenty-five days. This is exactly the contents of the second and the third columns of the table.

A geographical explanation

Let us have a look at the distances implied by the story. How far can one ride a horse on Earth without changing direction? In other words, what is the longest distance on land along one direction? Here are some examples:

- East Coast–West Coast distance in the US is approximately $3,000$ miles, or $4,828$ kilometers.
- Longest continuous distance on land along a longitude : $7,590$ kilometers (Northern Russia to Southern Thailand, $99°$ east).
- Longest continuous distance on land along a latitude: $10,726$ kilometers (Western France to Eastern China, $48°$ north.)
- Longest distance on land along any great circle: $13,573$ kilometers (Liberia to China).

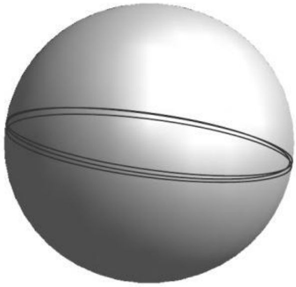
To find the distance covered so far by the prince and his knights (supposedly, they always move due south), we need to multiply 40 leagues by $3,120$ (approximately) days of travel. A league is an ancient measure of distance that varies from country to country, but is approximately equal to the distance a person can walk in an hour. We suppose that Dino Buzzati, being an Italian writer, would use the Roman league, equal to 1.4 mi, or 2.2225 km. In this case, the distance traveled by the prince at the time of the narration is $2.2225 \cdot 40 \cdot 3120 = 277,368$ km. No distance on land along one direction on Earth is that long, which suggests either a fantastic setting, or that the expedition, despite the hopes of the narrator, is not moving in the same direction. Let us choose middle ground: suppose that the expedition is taking place on Earth, but the arrangement of continents allows for an indefinite movement in one direction on land. Then we can imagine the path of the expedition as a tight spiral as in Figure 2.11 starting at the equator and circling around multiple times.

This model gives rise to another explanation for the discrepancy between the calculated time of Domenico's return ($3,125$ days) and the time reported by the narrator (between $3,118$ and $3,121$ days). The concept outlined below first came to the attention of astronomers in the Renaissance times, about the time the story of the seven messengers might be imagined to have occurred.

International date line

Recall the history of the first world circumnavigation headed by Ferdinand Magellan. He started to sail his five ships due west from the coast

Figure 2.11: Circling around the globe



of Spain on September 20, 1519. Almost three years later, on September 6, 1522, the eighteen survivors of his original crew (Magellan himself was killed in March 1521 in the Philippines) and his only surviving ship, *Victoria*, returned to Spain. But the ship's log had the arrival date marked as September 5, 1522. The log was recorded with utmost care and accuracy; in particular, the leap year 1520 was taken into account. The mystery of a missing day was discussed by the leading scientists of the time, among them the Venetian astronomer Gasparo Contarini, who suggested the correct explanation: moving westward, and sailing a whole circle around the Earth, you gain one day; moving eastward, you lose it. It was not until the nineteenth century when the International Date Line was established in the sense and position it has now: an imaginary line between the north and south poles at approximately 180° east separating Russia and Asia from the Americas, and one calendar day on Earth from the next. Now a person crossing the International Date Line traveling eastbound has to subtract a day; when traveling westbound, add a day.

If we suppose for a moment that the expedition was moving west-west-south instead of south (and was able to move continuously on land), the missing days can be explained by the same effect. A Renaissance setting of Buzzati's story suggests that the narrator (just like Magellan's crew) might not have known about the necessity to add a day for each complete circle when traveling westward on an Earth-like planet. Let us estimate the distances. The radius of Earth $R_\oplus$ is between $6,353$ and $6,384$ kilometers (larger at the equator). Therefore, the circumference at the equator is approximately $2\pi R_\oplus \simeq 40,090$ km. By the time of the narration, prince's expedition has covered 277, 368 kilometers. Dividing this distance by the circumference of the equator, we get $277,368/40,090 \simeq 6.9$. Therefore, six days (more than six if the narrator was moving along a higher latitude; less than six if he was deviating from the straight westward direction) might have to be added to the number of days in the narrator's log to obtain the actual number of days elapsed in the capital, or for a messenger who moves east and west, back and forth. The date discrepancy between a round-the-globe traveler and a stationary observer was one of the most remarkable successfully resolved mysteries of its time. Even if this argument was not intended by the author, it fits nicely with the story's Renaissance setting and its enigmatic character.

Messengers in space

With the advent of the Internet, the idea of sending a human messenger to deliver a letter might have lost some of its practical value. Nevertheless, there are real-world phenomena that essentially enact the story of the traveler and his messengers.

Can you think of a real-world situation where an object is moving away from an observer, and the messengers oscillate between them? A space probe such as *Voyager* is an example. This is an instrument, roughly a radio telescope, propelled by a rocket into deep outer space to send back to Earth information about remote planets and stars. The messengers are the photon particles that carry information between the space ship and the command center on Earth. There are multitudes of them, and their speed is the speed of light (in kilometers per second): $300,000 = 3 \cdot 10^5$ km/sec. The gravitational and relativistic effects have to be taken into account, which makes the trajectories curve and the time flow differently for different moving objects. The distances are much bigger, too: the probes *Voyager 1* and *Voyager 2*, launched in 1977, are now exploring the outermost layer of the heliosphere, the region of space dominated by the Sun. To give a rough idea of the scale of the distances, the approximate size of the Solar System is 4.5 billion kilometers ($4.5 \cdot 10^9$ kilometers). But the essence of the story of the messengers remains the same: *as the probe moves farther away from us, the messages we get from it become more and more outdated.* For example, on August 27, 2003, Mars came closest to the Earth in the previous 60,000 years. On that day a snapshot sent to Earth from the orbit of Mars took about 186 seconds (or 3.1 minutes) to arrive, while a snapshot sent from the edge of the Solar System communicates an observation made by the probe about 14, 700 seconds (about 4 hours) ago.

In fact, because the universe is expanding, any star is an object moving away from us, and its visible light is a messenger from it. Let us try to estimate how outdated our picture of the stars might be. The approximate diameter of our Galaxy is 100,000 light-years, or $9.4 \cdot 10^{17}$ km, and we are about 27,000 light-years away from the galactic center. This means that we see the stars at the opposite edge of our Galaxy the way they looked about 77,000 years ago. But we still receive their messengers, their beams of light, and maybe in another hundred thousand years they might be able to see the Solar System as it is now, and us in it.

EXERCISES

Question 2.1. James Bond is in Spyburg, 140 miles from an international border, and a villain is in Villainburg, 20 miles closer to the border along the same road. At noon, the villain starts driving toward the border at a constant speed of 80 miles per hour.

1. If James Bond drives at 100 miles per hour, when is the latest he should leave Spyburg to overtake the villain before he crosses the border?

2. If James Bond leaves Spyburg at 12:30 pm, what is the minimum average speed he has to maintain to overtake the villain before he crosses the border?

Question 2.2. Liz and Pat, who live 33 miles apart, want to go biking together. At 10 am they load their bikes in their cars and start driving toward each other's houses, Liz going at 50 miles per hour and Pat at 60 miles per hour. At the meeting point, they park their cars and immediately start riding their bikes at 18 miles per hour in the direction of Liz's house.

1. When will they arrive at Liz's house?

2. After spending an hour at Liz's house, they ride their bikes back to the parked cars. If Liz immediately starts driving home at 50 miles per hour, when will she arrive?

Question 2.3. Car A moves along a road at a constant speed of 60 miles per hour. Five miles ahead of it, car B moves in the same direction at a constant speed of 45 miles per hour. Thirty miles ahead of B, car C moves *in the opposite direction* at a constant speed of 55 miles per hour.

1. Which car will be first to meet B, A or C?

2. At the moment when B first meets with one of the cars, how far from it is the remaining car?

3. How much time will elapse between the meeting of car B with the first and the second cars?

Question 2.4. At 10 am car A starts moving east along a road at a constant speed of 50 miles per hour. At the same time 55 miles to the east of car A, car B starts moving west at 60 miles per hour, and car C starts moving east at 48 miles per hour. The moment cars A and B meet, B changes direction and starts moving east at 60 miles per hour.

1. When will cars B and C meet?

2. At the moment when B and C meet, how far behind is car A?

Question 2.5. In Example 2.3, suppose that Zoe starts from the eastern wall at 9:00:00, and Yolanda starts at the western wall at 9:00:05. The distance between the walls and the speeds of the bees are the same as before. Find the times of their first, second, and third meetings.

Question 2.6. In Example 2.3, suppose that Zoe starts from the eastern wall at 9:00:00, and Yolanda starts at the western wall at 9:00:05. The distance between the walls and the speeds of the bees are the same as before.

1. Determine where exactly between the walls the bees meet for the first, second, and third times.

2. Assuming that the bees fly forever, find all possible positions of their meetings.

Question 2.7. Starting from Pigeonville, a traveler walks at a constant speed along a straight road. After 1 hour of walking, he sends a trained pigeon with a letter back to Pigeonville. The pigeon flies at a constant speed of 8/5 times the speed of the traveler. Upon reaching Pigeonville, the pigeon immediately turns back to catch up with the traveler. When does the pigeon catch up with the traveler?

Question 2.8. Starting from Pigeonville, a traveler walks at a constant speed along a straight road. After 1 hour of walking, he sends a trained pigeon with a letter back to Pigeonville. The pigeon flies at a constant

Chapter 3

Acceleration: After the apple falls

While we can model constant speed motion by linear equations, we need more complicated quadratic equations to describe motion involving a constant *acceleration* – for example an apple falling from a tree. According to legend, this was the sight that inspired Sir Isaac Newton to write down his laws of motion and gravity, leading to the equations that govern the behavior of objects in the presence of a constant force.

In this chapter we will use some of the basic algebraic methods developed in Chapter 1 to describe accelerated motion of objects. The mathematical tool we will need to describe the position of an object moving with a constant acceleration is a quadratic equation. The acceleration may be caused by a force such as gravity for falling objects, or by braking force for vehicles slowing down on a road.

MATH

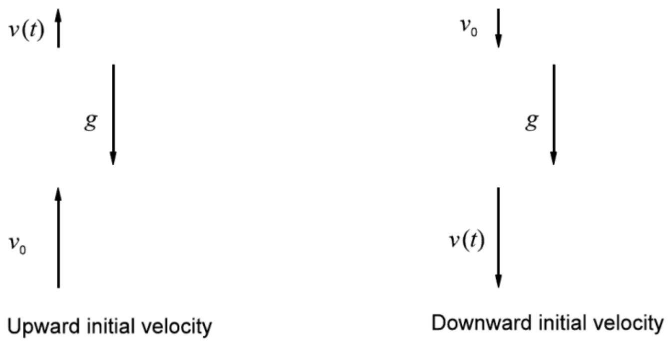
In Chapter 2, we considered objects moving at a constant speed along a straight line. This applies reasonably well to the motion of a particle (or an asteroid) in deep outer space, a ball rolling along a horizontal, frictionless surface, or a car in cruise control mode on a level freeway.

Now suppose we want to study the motion of a rock falling vertically downward. In this case, the velocity changes over time under the effect of the Earth's gravity. For the same reason, a rock thrown vertically upward begins to slow down until it reaches a maximum height, at which point its velocity reverses in direction and it starts to fall down, slowly increasing in speed as it accelerates to the ground. Some of the numerical characteristics of this familiar process may seem surprising. For instance, a rock thrown into the air with an initial velocity of 32 feet per second reaches the maximum height of 16 feet. If we double the initial velocity to 64 feet per second, the maximum height is 64 and not 32 feet, as one might expect.

The crucial observation to understanding this kind of motion is Newton's postulates that (1) the vertical velocity of a falling object changes linearly with time, and (2) the coefficient of this change, called the *gravitational acceleration*, is universal for all objects near the surface of the earth.

Let us write this statement in the form of an equation. Let v_0 be the initial (upward) vertical velocity of an object (if the velocity is directed downward, we will assume it is negative), and let $v(t)$ be its velocity at time t. Then the postulates state that for an object moving near the surface of the earth, the difference $v_0 - v(t)$ is proportional to the time interval t with the coefficient given by the gravitational acceleration.

Figure 3.1



We have

$$v(t) - v_0 = -gt,$$

where

$$-g \simeq -9.8 \text{ m/sec}^2 \simeq -32 \text{ ft/sec}^2$$

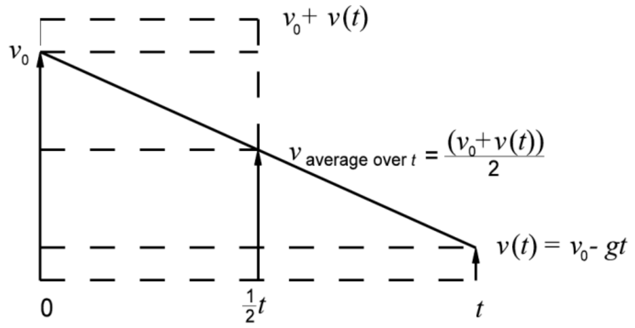
is the (downward) gravitational acceleration on Earth.

Equivalently, we have the following formula relating the vertical velocity of an object with time:

$$v(t) = v_0 - gt.$$

We would like to know the position of the object (along a vertical line) whose velocity is given by this expression. We cannot multiply the velocity $v(t)$ by time as we did in Chapter 2, because now the velocity is not constant. However, we can find the *average* velocity over time t. Because the velocity is changing linearly, the average velocity is half the sum of the initial and the final velocities, as is shown in the diagram below. It also equals the velocity at half-time $\frac{1}{2}t$.

Figure 3.2



$$v_{\text{average over } t} = \frac{1}{2}(v_0 + v(t)) = \frac{1}{2}(v_0 + v_0 - gt) = v_0 - \frac{gt}{2}.$$

Now we can compute the position of the object as it depends on time. Let s_0 be the initial position (measured upward from a certain level, for example, from the ground):

$$s(t) = s_0 + (v_{\text{average over } t})t = s_0 + v_0 t - \frac{gt^2}{2}.$$

Along with the initial postulate that the acceleration $a(t)$ is constant, we obtain a familiar set of equations, often called the *equations of motion*:

$$\begin{aligned} a(t) &= -g, \\ v(t) &= v_0 - gt, \\ s(t) &= s_0 + v_0 t - \frac{1}{2} g t^2. \end{aligned}$$

These equations allow us to find the position at any moment of time of an object moving with constant acceleration, its initial velocity, or how long it takes for the object to reach any given elevation. In the latter case, we may need to solve a quadratic equation.

While working with these equations, we have to pay attention to the physical dimensions of the acceleration $a(t)$, velocity $v(t)$, and displacement $s(t)$. They should all be expressed either in the metric (m, m/sec, m/sec^2) or in the English (ft, ft/sec, ft/sec^2) system of measure.

Example 3.1. A rock is thrown vertically upward from an initial elevation of 1 meter with an initial velocity of 4 meters per second. How long will it take before the rock falls on the ground?

Solution: Denote by t_{final} the moment of time when the rock hits the ground. At this moment, the vertical position of the rock is $s(t_{\text{final}}) = 0$. Therefore, we have an equation for t_{final}:

$$0 = 1 + 4t_{\text{final}} - \frac{1}{2} 9.8 \cdot t_{\text{final}}^2 \quad \Longrightarrow \quad -4.9 t_{\text{final}}^2 + 4 t_{\text{final}} + 1 = 0.$$

To solve this equation for t_{final}, we apply the formula for a general solution of a quadratic equation derived in Chapter 1:

$$ax^2 + bx + c = 0, \quad a \neq 0 \quad \Longrightarrow \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Substituting $a = -4.9$, $b = 4$, and $c = 1$, we obtain:

$$t_{\text{final}} = \frac{-4 \pm \sqrt{4^2 - 4 \cdot (-4.9) \cdot 1}}{2 \cdot (-4.9)} \simeq \frac{-4 \pm 6}{-9.8} \simeq 1 \text{ sec.}$$

Here we have discarded the negative answer. It will take the rock about 1 second to fall to the ground. □

The equations of motion are applicable in a wide range of problems. For instance, they describe the vertical motion of objects under the influence of any gravitational force, with the value of the gravitational acceleration

g determined according to the given situation (see Examples 3.12 and 3.13 for the vertical motion of objects on the Moon).

More generally, the same equations describe the motion of an object along a straight line with a constant acceleration caused by any external force. In this case, we have to replace g by the acceleration a determined by the given force.

APPLICATIONS

In the rest of this chapter, we will apply the equations of motion to objects moving under some form of constant force.

Vertical motion near the surface of the Earth

The problems on the motion of objects influenced by the Earth's gravity may involve solving equations for the initial velocity, or some measure of time, or the position of the object at a given moment of time. We start with the example given at the beginning of the chapter.

Example 3.2. If you throw a rock upward from the ground with the initial velocity 32 feet per second, it will reach a maximum height of 16 feet. What is the maximum height attained by a rock thrown upward with an initial velocity 64 feet per second?

Solution: Contrary to a possible first guess, it is not 32 feet. Let us express the maximum height reached by the rock in terms of its initial velocity. At the instant when the rock is at its highest, its velocity is zero because it has just finished traveling upward and is about to start falling under gravity. Thus, we first solve for the time t_{max} at which the rock reaches the maximum height given that $v(t_{\text{max}}) = 0$:

$$v(t_{\text{max}}) = v_0 - gt_{\text{max}} = 0 \qquad \Longrightarrow \qquad t_{\text{max}} = \frac{v_0}{g}.$$

The elevation of the rock at t_{max} is

$$s(t_{\text{max}}) = 0 + v_0 t_{\text{max}} - \frac{1}{2} g t_{\text{max}}^2 = v_0 \frac{v_0}{g} - \frac{1}{2} g \left(\frac{v_0}{g}\right)^2 = \frac{v_0^2}{g} - \frac{1}{2}\frac{v_0^2}{g} = \frac{1}{2}\frac{v_0^2}{g}.$$

We conclude that the highest elevation s_{max} reached by an object thrown from the ground with an initial velocity v_0 is

$$s_{\text{max}} = s(t_{\text{max}}) = \frac{1}{2}\frac{v_0^2}{g}.$$

Example 3.5. A cannonball is fired upward from a toy cannon on the ground, and it reaches a height of 32 feet in 2 seconds. Find (a) the initial velocity, and (b) the times at which it reaches 20 feet.

Solution: For this problem we have $s_0 = 0$ ft, while v_0 is unknown.
(a) We use the equation for the position $s(t)$, with given time $t = 2\,\text{sec}$ and position $s(t) = s(2) = 32\,\text{ft}$. Thus,

$$32 = s(2) = 0 + v_0 \cdot 2 - 16 \cdot 2^2,$$

which simplifies to: $2v_0 = 32 + 64 = 96\,\text{ft/sec}$. Hence $v_0 = 48$ ft/sec.
(b) Having determined $v_0 = 48$ ft/sec, we now use the same equation to solve for the unknown time(s) t at which $s(t) = 20$ ft. We compute:

$$20 = 0 + 48t - 16t^2 \qquad \Longrightarrow \qquad 16t^2 - 48t + 20 = 0.$$

We can now solve for the roots of the quadratic equation, or first divide all terms by 16 in order to work with smaller numbers. If we do so, the equation reduces to: $t^2 - 3t + \frac{5}{4} = 0$. Now apply the formula for the roots, with $a = 1, b = -3, c = \frac{5}{4}$, to obtain:

$$t = \frac{-(-3) \pm \sqrt{(-3)^2 - 4 \cdot 1 \cdot (5/4)}}{2} = \frac{3 \pm \sqrt{4}}{2} = 0.5, \quad 2.5 \text{ sec}.$$

Thus, the cannonball is 20 feet high at $t = 0.5$ sec and $t = 2.5$ sec. □

Bouncing balls

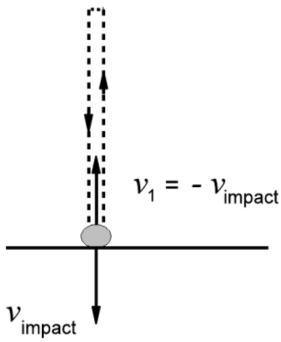
Suppose you drop a tennis ball from a certain height to the ground. How high will it bounce? Let us consider a couple of cases.

Example 3.6. Imagine a rubber ball that bounces without any energy loss, so that its velocity right after it hits the ground is the negative of its velocity right before it hits the ground. If such a ball is dropped down from height s_0, how high will it bounce back?

Solution: First we have to find the time t_{impact} it takes for the ball to fall, and its velocity at the impact v_{impact}. With the initial velocity zero, the position equation gives

$$s(t_{\text{impact}}) = s_0 - \frac{1}{2} g t_{\text{impact}}^2 = 0,$$

Figure 3.3



because at time t_{impact} the ball is on the ground. Then solving for v_{impact} gives

$$\frac{1}{2}gt_{\text{impact}}^2 = s_0, \quad \Longrightarrow \quad t_{\text{final}} = \sqrt{\frac{2s_0}{g}},$$

where we have discarded the negative solution. Because the initial velocity was zero, the velocity at time t_{impact} is

$$v_{\text{impact}} = v(t_{\text{impact}}) = 0 - gt_{\text{impact}} = -g\sqrt{\frac{2s_0}{g}} = -\sqrt{2gs_0}.$$

By assumption, the velocity v_1 after the ball hits the ground is the negative of v_{impact}. This is the initial velocity for the bounce:

$$v_1 = -v_{\text{impact}} = \sqrt{2gs_0}.$$

Now we can use the formula obtained in Example 3.2 for the maximum height attained by the ball with a given initial velocity:

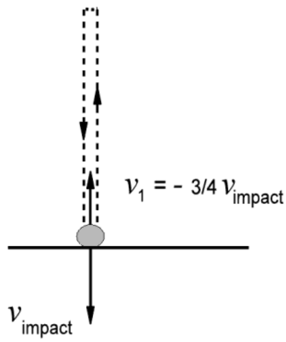
$$s_{\max} = \frac{1}{2} \cdot \frac{v_1^2}{g} = \frac{1}{2} \cdot \frac{2gs_0}{g} = s_0.$$

The ball will bounce back to the exact same height s_0. □

This result may look surprising to you, and for a good reason: in reality any rubber ball loses energy, and therefore velocity, as it hits the ground. Here is a more realistic example.

Example 3.7. A tennis ball is dropped from a height of 8 feet. Each time it bounces off the ground, its vertical velocity reverses direction and loses one quarter of its magnitude. Find the maximum height attained by the ball on the first and the second bounce.

Figure 3.4



Solution: The downward velocity v_{impact} before the ball hits the ground for the first time is determined by the same formula as in the previous example, $v_{\text{impact}} = -\sqrt{2gs_0}$. Then the upward velocity for the first bounce v_1 is three quarters of it, taken with the positive sign:

$$v_1 = \frac{3}{4}\sqrt{2gs_0}.$$

This is the initial velocity for the first bounce. The maximum height s_1 attained at the first bounce with the initial velocity v_1 is

$$s_1 = \frac{1}{2}\frac{v_1^2}{g} = \frac{1}{2}\frac{(\frac{3}{4})^2 \cdot 2 \cdot g \cdot s_0}{g} = \left(\frac{3}{4}\right)^2 \cdot s_0 = \frac{9}{16} \cdot 8 = 4.5 \text{ ft}.$$

To find the height attained on the second bounce, we have to repeat the above computation with the initial height of the drop $s_1 = 4.5$ ft. We can avoid going through the computation by noticing that the maximum height of the first bounce s_1 is related to the initial height s_0 by the formula

$$s_1 = \left(\frac{3}{4}\right)^2 \cdot s_0.$$

Because we are solving the exact same question with s_0 replaced by s_1, the maximum height of the second bounce s_2 is given by the formula

$$s_2 = \left(\frac{3}{4}\right)^2 \cdot s_1 = \left(\frac{3}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 \cdot s_0 = \left(\frac{3}{4}\right)^4 \cdot s_0 = \frac{81}{256} \cdot 8 = \frac{81}{32} \simeq 2.53 \text{ ft.}$$

Now we could easily compute the maximum height attained in any subsequent bounce of the ball: each time it decreases by a factor of $(\frac{3}{4})^2$. □

Practice problem 3.8. In the conditions of Example 3.7, what is the time interval between (a) the first and the second bounce of the ball, (b) the second and the third bounce?
Answer: (a) $t_1 = \frac{3}{2}\sqrt{\frac{2s_0}{g}} \simeq 1.06\,\text{sec}$, (b) $t_2 = \frac{3}{4}t_1 \simeq 0.8\,\text{sec}$.

Examples with more than one moving object

Systems with more than one object moving under the influence of gravity lead to more complicated models. For example, you can consider two balls launched from different heights with different initial velocities. Newton's equations of motion allow us to find moments of time and positions where the balls might meet.

Example 3.9. A yellow tennis ball is dropped from a height of 5 meters, and at the same moment of time a white tennis ball is thrown upward from the ground. Find the initial upward velocity of the white ball if the balls meet exactly halfway, at a height of 2.5 meters above the ground.

Solution: Let us denote by $s_{\text{y}}(t)$ and $s_{\text{w}}(t)$ the position of the yellow and the white balls, respectively, as they change with time. Then we have

$$s_{\text{y}}(t) = s_0 - \frac{1}{2}gt^2, \qquad s_{\text{w}}(t) = v_0 t - \frac{1}{2}gt^2,$$

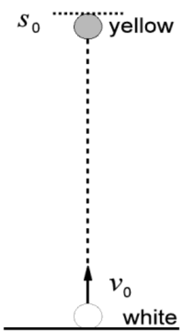
where $s_0 = 5$ m is the initial height of the yellow ball, and v_0 is the unknown initial velocity of the white one. At the moment t_{meet} of the meeting the balls have the same vertical position:

$$s_0 - \frac{1}{2}gt_{\text{meet}}^2 = v_0 t_{\text{meet}} - \frac{1}{2}gt_{\text{meet}}^2.$$

Then we have

$$s_0 = v_0 t_{\text{meet}} \quad \Longrightarrow \quad t_{\text{meet}} = \frac{s_0}{v_0}.$$

Figure 3.5



This formula is familiar from the constant speed motion model. It implies that the time before the balls meet is the same as it would be in the absence of gravity. This happens because gravity affects both balls in the same way, so that their *relative* position is independent of it. But the position of their meeting s_{meet} with respect to the ground level, of course, depends on g:

$$s_{\text{meet}} = s_{\text{y}}(t_{\text{meet}}) = s_0 - \frac{1}{2}gt_{\text{meet}}^2 = s_0 - \frac{g}{2}\cdot\left(\frac{s_0}{v_0}\right)^2.$$

We are given that this height is half of s_0. Then

$$s_{\text{meet}} = \frac{1}{2}s_0 = s_0 - \frac{g}{2}\cdot\left(\frac{s_0}{v_0}\right)^2 \quad\Longrightarrow\quad \frac{g}{2}\cdot\left(\frac{s_0}{v_0}\right)^2 = \frac{1}{2}s_0 \quad\Longrightarrow\quad \frac{s_0}{v_0^2} = \frac{1}{g}.$$

This leads to

$$v_0^2 = s_0 \cdot g \quad\Longrightarrow\quad v_0 = \sqrt{s_0 \cdot g} = \sqrt{5\cdot 9.8} = \sqrt{49} = 7\,\text{m/sec},$$

where we have discarded the negative answer for the velocity. The white ball was launched upward with a velocity of 7 m/sec. □

Example 3.10. A black rubber ball is thrown upward from the ground with an initial velocity $v_0 = 12\,\text{ft/sec}$. At the same time, a white rubber ball is thrown upward from a height of 2 feet with an initial velocity $\frac{1}{2}\,v_0 = 6\,\text{ft/sec}$. When and at what height will the balls meet?

Solving for v_0, we obtain $v_0^2 = 19.6 \cdot 140 = 2,744$. Therefore $v_0 \simeq 52.4$ m/sec. □

According to Wikipedia, the speed of water at the nozzle in the Jet d'Eau is 200 kilometers per hour, which is approximately 55.6 meters per second. Indeed, in reality the speed must be higher than our estimation because of the air resistance to the motion of the water (which we had ignored).

Soccer on the Moon

The strength of Newton's law of gravitational attraction, and hence the equations of motion, lies in their generality. In particular, any celestial body exerts gravitational force, and an object placed at a given distance from the source of gravity will experience a constant acceleration that is the same for all objects and depends only on the mass of the source of gravity. Here is an amusing, if not very realistic, example.

Example 3.12. In the animated film *A Grand Day Out*, Wallace and Gromit are having a picnic on the Moon. Wallace kicks a soccer ball and waits 7 seconds for it to fall. When the ball does not come back in 7 seconds, Wallace walks on. How long should he have waited for the ball to return?

Solution: We need some additional input. First, we need to know the value of the gravitational acceleration on the surface of the Moon. Online sources provide the number $g_{\text{moon}} = 1.6\,\text{m/sec}^2$. Second, we don't know exactly how hard Wallace kicked the ball. You can try emailing Wallace or his creators, but we found it easier first to look for estimates of how fast a soccer ball can be kicked in theory. We quickly found that Lisbon's left-back Ronny Heberson holds the record of nearly 132 miles per hour, or 59 meters per second, and that the average for professionals is about half as fast. Now, Wallace is no professional; in fact, we suspect him to be more of a spectator than a participant when it comes to sports. We won't be too far off if we assume he kicked the ball at about 15 meters per second. Now we have all the ingredients to estimate the time before the ball should return to the surface (of the Moon!).

We can take the height of the kick to be the zero level, $s_0 = 0$. We want to find the time t it takes the ball to return to the same level: $s(t) = 0$. According to the equations of motion,

$$0 = 0 + v_0 t - \frac{g_{\text{moon}} t^2}{2},$$

where $v_0 = 15\,\text{m/sec}$, and $g_{\text{moon}} = 1.6\,\text{m/sec}^2$. Solving for t, we have

$$0 = t\left(v_0 - \frac{g_{\text{moon}} t}{2}\right) \qquad \Longrightarrow \qquad \left[\begin{array}{l} t = 0; \\ v_0 - \frac{gt}{2} = 0. \end{array}\right.$$

The first solution, $t = 0$, corresponds to the moment of the kick. We are interested in the second solution given by the equation

$$v_0 = \frac{g_{\text{moon}} t}{2} \qquad \Longrightarrow \qquad 2v_0 = g_{\text{moon}} t \qquad \Longrightarrow \qquad \frac{2v_0}{g_{\text{moon}}} = t.$$

Plugging in the numbers, we get

$$t = \frac{2v_0}{g_{\text{moon}}} = \frac{2 \cdot 15\,\text{m/sec}}{1.6\,\text{m/sec}^2} = 19\ \text{sec}.$$

Wallace should have waited for 19 seconds.

For comparison, let us see how long it would take for the ball kicked with the same initial velocity to fall to Earth. The equation is the same; the only difference is in the value of g – on Earth it is 9.8 meters per second squared. We have

$$t = \frac{2v_0}{g} = \frac{2 \cdot 15\,\text{m/sec}}{9.8\,\text{m/sec}^2} \simeq 3\ \text{sec}.$$

No wonder Wallace got impatient – he waited for more than twice the time the ball would take to fall down to Earth! (Whereas he should have waited for about *six* times the time.) □

We can also compute the maximum height the soccer ball will reach on the Moon.

Example 3.13. Wallace kicked a soccer ball from the surface of the Moon upward with an initial velocity of 15 meters per second. What is the maximum height it will reach? If Ronny Heberson kicked the ball instead, how high would it go?

Solution: For any object under the influence of gravity that is thrown upward with a given initial velocity, the formula for the maximum height was derived in Example 3.2:

$$s_{\max} = \frac{1}{2} \cdot \frac{v_0^2}{g}.$$

On the Moon we have $g_{\text{moon}} = 1.6\,\text{m/sec}^2$, and the initial velocity is given to be $v_0 = 15\,\text{m/sec}$. Then we have

$$s_{\max} = \frac{1}{2} \cdot \frac{15^2}{1.6} \simeq 70.3 \text{ m}.$$

So Wallace kicked the ball to a height of about 70 meters.

Now, let us assume that Ronny Heberson would kick the ball at about 60 meters per second, which is four times the initial velocity of Wallace's kick, $v_{\text{Heberson}} = 4v_0$.

Because the maximum height given by the formula is proportional to the square of the velocity, it will increase by a factor of 16:

$$s_{\text{Heberson}} = \frac{1}{2} \cdot \frac{(4v_0)^2}{g_{\text{moon}}} = 16\ s_{\max} \simeq 16 \cdot 70.3 \simeq 1,125 \text{ m}.$$

The ball would go upward more than one kilometer! □

Braking distance on a highway

The same laws and equations of motion that apply to objects falling under gravity hold when you are driving on a freeway and apply the brakes in order to stop. Namely, it is reasonable to assume that the action of the brakes provides a constant negative acceleration $a(t) = -a$, which may depend on the vehicle in question. If we denote the velocity at the instant you start braking as v_0, and measure time starting from that same instant, then the equations describing the position and velocity of the vehicle are:

$$a(t) = -a, \qquad v(t) = v_0 - at, \qquad s(t) = s_0 + v_0 t - \frac{1}{2}at^2.$$

This allows us to perform similar calculations as in the above examples, to model the motion of a car braking on a freeway.

Example 3.14 (Braking distance)**.** Suppose that if you are driving at 30 miles per hour, and applying the brakes hard, your car comes to a stop in 500 feet. If instead the same car is speeding at 60 miles per hour, what is the stopping distance upon applying the brakes?

Solution: To solve this problem, we need to carry out the same analysis as in Example 3.2. However, first we have to convert all quantities into the same physical units. There are 5, 280 feet in a mile and 3, 600 seconds in an hour. Then

$$30 \text{ mph} = 30 \cdot \frac{5,280}{3,600} \text{ ft/sec} = 44 \text{ ft/sec}.$$

Similarly, 60 mph = 88 ft/sec.

Now suppose $SD(v_0)$ denotes the stopping distance for the car, from the point where the brakes were applied and the car had velocity v_0 ft/sec. At the moment when the car stops, its velocity is zero:

$$v(t_{\text{stop}}) = v_0 - at_{\text{stop}} = 0 \quad \Longrightarrow \quad t_{\text{stop}} = \frac{v_0}{a}.$$

Plugging t_{stop} in the equation for $s(t)$, just like in Example 3.2, we obtain the stopping distance:

$$SD(v_0) = \frac{1}{2} \cdot \frac{v_0^2}{a_{\text{car}}},$$

where $-a_{\text{car}}$ is the negative acceleration of the car in feet per second squared, caused by braking. Given the stopping distance at the initial velocity of 30 miles per hour, we can determine a_{car}:

$$500 = SD(44) = \frac{44^2}{2a_{\text{car}}} = \frac{968}{a_{\text{car}}}.$$

This yields: $a_{\text{car}} = 968/500 = 1.936$ ft/sec^2. Now plugging in the initial velocity 60 mph = 88 ft/sec, we compute:

$$SD(88) = \frac{88^2}{2a_{\text{car}}} = \frac{7,744}{3.782} = 2,000 \text{ ft}.$$

Thus, just like in Example 3.2, the stopping distance quadruples when the initial speed doubles. □

The conclusion we can draw from the previous example is: the faster we are driving, the more distance we will need to brake to a complete stop.

EXERCISES

Question 3.1. A passenger in a hot air balloon which is stationary at a height of 80 feet above the ground drops a stone. Compute the time and velocity at which (a) it is 40 feet above the ground, and (b) it hits the ground. Is the time in part (b) twice of the time in part (a), or more than twice, or less? Once you have computed the two times, how do you explain the answer to the previous question?

Question 3.2. A person shoots an arrow vertically up from the ground with some initial velocity, and it reaches its maximum height in 3.5 seconds. Compute (a) the initial velocity, (b) the maximum height attained, and (c) the time(s) in which it reaches half of this height.

Question 3.3. A woman throws a tennis ball down from a height of 5 meters with an initial downward velocity of 5 meters per second. Each time it bounces off the ground, its upward velocity is four-fifths of the downward velocity at the moment of impact. Find the maximum height attained by the ball on the first and the second bounce.

Question 3.4. Venus Williams drops a tennis ball from a height of 6 meters, and at the same time Serena Williams throws another tennis ball upward from the ground with an initial velocity of 9 meters per second. Where and when will the balls meet?

Question 3.5. Solve Example 3.9 with the assumption that the balls meet at seven-eighths of the initial height of the yellow ball.

Question 3.6. Anna throws a small snowball vertically upward from the ground with an initial velocity v_0. At the same time, Elsa throws another snowball from a height $s_0 = 4\,\mathrm{m}$ above the ground with an upward initial velocity $\frac{1}{2}\,v_0$.

1. Find v_0 if the snowballs meet at the height $s_0 = 4\,\mathrm{m}$.

2. Find the velocities of both snowballs at the time they meet.

Question 3.7. A British officer fires a cannonball vertically upward from a cannon located on the ground. It subsequently bursts through a barrier which is placed at a certain height s_0 above the ground. Upon bursting through the barrier, the velocity of the cannonball slows down to 150 feet per second. Eight seconds after bursting through the barrier, the cannonball is 226 feet high. Compute the height of the barrier above the ground. Also compute all time(s) when the cannonball is 226 feet high, and its velocities at those times.

Question 3.8. Suppose you are at a window located 70 feet above the ground, and a pickup truck is coming along the street toward you at a constant speed of 20 miles per hour. You want to drop a small object into the trunk of the truck. At the moment when you drop it, how far should the truck be from the point directly under your window?

Question 3.9. The height of Niagara Falls is about 165 feet. Estimate the vertical velocity of the water stream at the bottom.

Note that in the second case we are unable to express the answer without using the square root sign. The number $\sqrt{5}$ is nothing but the name given to the positive solution of this particular equation.

A real number that can be expressed as a quotient of two integers is called *rational*. All other real numbers are called *irrational*. For example, $\frac{7}{6}$ is rational and $\sqrt{5}$ is irrational. Another example of an irrational number is $\pi = 3.14159265\ldots$. The decimal portion of π is infinitely long and never repeats itself. For the number to be irrational, its decimal expression needs to be infinitely long and non-periodic. We will show in Chapter 10 that any decimal number with a periodically repeating "tail" is rational.

Example* 4.3. How can we be sure that $\sqrt{5}$ is not a quotient of two integers?

Solution: Let us suppose for a moment that there are integers p and q such that $\sqrt{5} = \frac{p}{q}$, and the denominator q is the smallest positive integer with this property. (This means we use the fraction $\frac{7}{3}$ instead of $\frac{14}{6}$). Multiplying by q and taking the square of both sides gives

$$\sqrt{5}q = p \quad \Longrightarrow \quad 5q^2 = p^2.$$

Because p and q are integers, this means that p^2 is divisible by 5. But then p itself is also divisible by 5. An example that satisfies this condition is 100, which is divisible by 5, but then its square root 10 is also divisible by 5. If we take 4 instead of 5, then 36 is a square and divisible by 4, but 6 is not. The reason is that $4 = 2^2$ is a square of an integer, and 5 *is not*. So, in our case, because p^2 is divisible by 5, we can always write $p = 5k$ for some new integer k. Now our equality reads

$$5q^2 = (5k)^2 \quad \Longrightarrow \quad 5q^2 = 25k^2 \quad \Longrightarrow \quad q^2 = 5k^2,$$

where q and k are integers. This is just like the equation we had before for p and q. By the same argument, q^2 is divisible by 5, and so is q. Let $q = 5m$ for an integer m. Then

$$\sqrt{5} = \frac{p}{q} = \frac{5k}{5m} = \frac{k}{m},$$

with m a positive integer smaller than q, which contradicts the assumption that q was the smallest positive denominator of a fraction equal to $\sqrt{5}$. Therefore, the number $\sqrt{5}$ is irrational. □

Examples of irrational numbers include square roots of integers that are not complete squares, for instance $\sqrt{2}$ or $\sqrt{12}$, cube roots of integers that are not cubes, like $\sqrt[3]{7}$, and so on. Multiplying an irrational number by a rational coefficient or adding a rational number to it produces again an irrational number, as the next example shows.

Example 4.4. Is the number $\frac{7\sqrt{2}}{10} + 3$ rational or irrational?

Solution: Suppose that $a = \frac{7\sqrt{2}}{10} + 3$ is rational and equals $\frac{p}{q}$, where p and q are integers. Then $a - 3 = \frac{p}{q} - 3 = \frac{p-3q}{q}$ is rational. So $\frac{10}{7}(a-3) = \frac{10}{7} \cdot \frac{p-3q}{q} = \frac{10p-30q}{7q}$ is also rational. But $\frac{10}{7}(a-3) = \sqrt{2}$, which we know is irrational, a contradiction. Therefore, a is irrational. □

The golden ratio

Here is a quadratic equation whose irrational solution is quite famous. Suppose you want to divide a segment of a line into two parts, so that the ratio of the larger part (a units long) to the smaller part (b units long) is the same as the ratio of the whole (c units long) to the larger part. This is expressed in the equation:

$$\frac{a}{b} = \frac{c}{a}, \quad \text{or} \quad \frac{a}{b} = \frac{a+b}{a}, \quad \text{or} \quad \frac{a}{b} = 1 + \frac{b}{a}.$$

With the notation $\phi = \frac{a}{b}$, the equation becomes: $\phi = 1 + \frac{1}{\phi}$. Multiplying by ϕ, we get the quadratic equation $\phi^2 = \phi + 1$, or

$$\phi^2 - \phi - 1 = 0, \tag{4.1}$$

whose solutions are $\phi_1 = \frac{1+\sqrt{5}}{2}$, $\phi_2 = \frac{1-\sqrt{5}}{2}$. We are interested in the positive solution:

$$\frac{a}{b} = \phi = \frac{1+\sqrt{5}}{2} \simeq 1.6180339887.$$

The number ϕ is irrational by an argument similar to Example 4.4, and is called the *golden ratio*, or the golden mean. It exhibits many amazing properties. We will start by pointing out its relation to the Fibonacci sequence.

The *Fibonacci sequence* $\{f_n\}$ is the sequence of integers starting with $1, 1$, and such that each next element of the sequence is the sum of two previous elements, $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 3$:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ \ldots$$

Thus, $f_1 = 1$, $f_2 = 1$, $f_3 = f_1 + f_2 = 2$, $\quad \ldots, \quad f_6 = f_5 + f_4 = 5 + 3 = 8 \ldots$

Let us look at the sequence of ratios of consecutive Fibonacci numbers $\frac{f_{n+1}}{f_n}$:

$$\frac{1}{1} = 1,\ \frac{2}{1} = 2,\ \frac{3}{2} = 1.5,\ \frac{5}{3} \simeq 1.667,\ \frac{8}{5} = 1.6,\ \frac{13}{8} = 1.625,\ \frac{21}{13} \simeq 1.615 \ldots$$

As we go further along the Fibonacci sequence, the ratios become closer to each other and seem to approach a certain limit number. This number can only be the golden ratio. Here is why: consider three consecutive numbers $b, a, b + a$ far along in the Fibonacci sequence. Then the ratios $\frac{a}{b}$ and $\frac{b+a}{a}$ should be very close to the limit number, and therefore almost equal. But $\frac{a}{b} = \frac{b+a}{a}$ is the equation for the golden ratio. Indeed, the ratios of the form $\frac{f_{n+1}}{f_n}$ provide a good approximation for ϕ. For example, $\frac{f_{20}}{f_{19}} = \frac{6{,}765}{4{,}181} \simeq 1.6180339632$. We will discuss other properties of ϕ in Application 2.

APPLICATION 1

Do irrational numbers have any practical use? One unexpected application is in paper manufacturing. Suppose you want to produce sheets of paper in a variety of sizes (for posters, letters, memos). To minimize production costs, it would be good if a large sheet of paper could be cut up into a number of smaller sheets. Besides, you would like for all sizes to have the same aspect ratio (length to width), for convenient enlargement and reduction of pages. It turns out that the two requirements together lead to a quadratic equation for the aspect ratio, whose solution is an irrational number.

The Lichtenberg conditions

To be more specific, the international paper size standard ISO 216 defines the series An for $n = 0, 1, 2, \ldots 10$ by the following requirements:

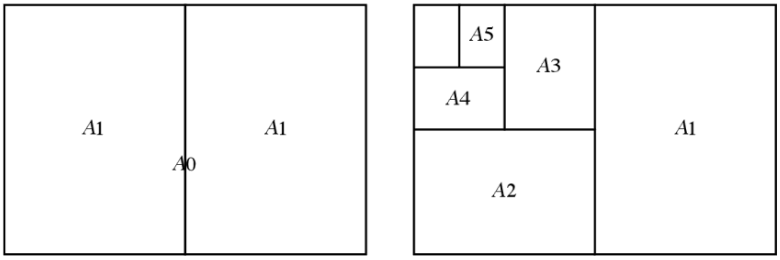
1. A landscape-oriented sheet of paper of size An, when cut in half, produces two portrait-oriented sheets of paper of size $A(n + 1)$ for all $n = 0, 1, 2, \ldots 9$.

2. The aspect ratio for all sizes is the same.

3. The area of a sheet of paper of size $A0$ is 1 square meter.

The idea of introducing conditions (1) and (2) to ensure production efficiency and convenient resizing goes back to the German scientist Georg Christoph Lichtenberg, who proposed it at the end of the eighteenth century.

Figure 4.1: The An paper sizes.



$A0$ divided into two $A1$s $A0$ to $A5$ paper sizes

First let us find the aspect ratio of the An sizes. Let a denote the length and b the width of a size An. Then by (1), b and $\frac{a}{2}$ are, respectively, the length and width of the size $A(n+1)$. To have the same aspect ratio, the parameters should satisfy the equation

$$\frac{b}{a} = \frac{a/2}{b}.$$

Multiplying by b and a, we get

$$b^2 = \frac{a^2}{2} \quad \Longrightarrow \quad 2b^2 = a^2 \quad \Longrightarrow \quad a = \sqrt{2}b.$$

Thus, $\sqrt{2}$ is the aspect ratio common to all An sizes.

Example 4.5. What is the ratio between the length of the $A4$ size and the length of the $A0$ size?

Solution: Let us denote by $l0, l1, l2, l3, l4$ the lengths of the $A0, A1, A2, A3$, $A4$ sizes, respectively. Then $l1$ is the width of the $A0$ size, which equals $l0/\sqrt{2}$. Similarly, $l2 = l1/\sqrt{2}$, $l3 = l2/\sqrt{2}$, and $l4 = l3/\sqrt{2}$. Taking all this into account, we get

$$l4 = \frac{l3}{\sqrt{2}} = \frac{l2}{\sqrt{2}\cdot\sqrt{2}} = \frac{l2}{2} = \frac{l1}{2\cdot\sqrt{2}} = \frac{l0}{2\cdot\sqrt{2}\cdot\sqrt{2}} = \frac{l0}{4}.$$

Finally, $\frac{l4}{l0} = \frac{1}{4}$. □

Now let us find the dimensions of the $A0$ size. If a is the length and b the width, then $a \cdot b = 1m^2$ and $a/b = \sqrt{2}$. We have

$$ab = a\frac{a}{\sqrt{2}} = \frac{a^2}{\sqrt{2}} = 1.$$

Therefore, $a^2 = \sqrt{2}$ and $a = \sqrt{\sqrt{2}}$. What kind of a number is this? Clearly, $\sqrt{\sqrt{2}} \cdot \sqrt{\sqrt{2}} = \sqrt{2}$. Also, $\sqrt{2} \cdot \sqrt{2} = 2$. Therefore, $\sqrt{\sqrt{2}}$ multiplied by itself *four* times gives 2. This irrational number is called the 4th root of 2, and denoted by $\sqrt[4]{2}$. Using your calculator, you can find $\sqrt[4]{2} = \sqrt{\sqrt{2}} \simeq 1.189207$. For the width b we find: $b = \frac{1}{a} = \frac{1}{\sqrt[4]{2}} \simeq 0.840896$. In fact, the dimensions of the $A0$ size are, in millimeters, $1,189\,\text{mm} \times 841\,\text{mm}$.

We will talk more about roots of various degrees in Chapter 5.

Example 4.6. Find the length and the area of the $A5$ size.

Solution: Proceeding just as in Example 4.5, we find that the ratio of the length $l5$ of the size $A5$ to the length $l0$ of the size $A0$ is $l5/l0 = \frac{1}{4\sqrt{2}}$. Therefore, $l5 = l0/(4\sqrt{2}) = \frac{\sqrt[4]{2}}{4\sqrt{2}} \simeq 0.210224...$ In fact, ISO 216 lists the length of $A5$ to be equal to 210 millimeters. To find the area, recall the condition (1): each smaller size has half the area of a larger size. For the area of the $A5$ size we have:

$$A5 = \frac{1}{2}A4 = \frac{1}{4}A3 = \frac{1}{8}A2 = \frac{1}{16}A1 = \frac{1}{32}A0 = \frac{1}{32}\,\text{m}^2 = 0.03125\,\text{m}^2.$$

□

Other standard paper sizes

In addition to the An sizes, there are two more series: Bn sizes and Cn sizes. The Bn sizes are determined by the conditions:

(1b) The length of the Bn size is the *geometric mean*, or the square root, of the product of the lengths of the $A(n-1)$ and An sizes:

$$l(Bn) = \sqrt{l(A(n-1)) \cdot l(An)},$$

and $l(B0) = \sqrt{2} \cdot l(B1)$.

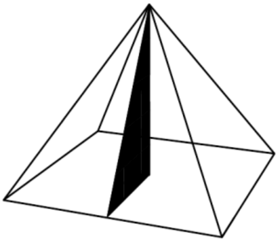
Figure 4.2: Great Pyramid of Giza (photo by Jerome Bon).

Figure 4.3: Parthenon (left, photo by Tilemahos Efthimiadis), and Athena, small replica of a statue by Phidias (right, photo by William Neuheisel).

regular pentagonal faces is the major element of the composition. (For the relation of a regular pentagon and the golden ratio, see Practice Problem 4.9.)

Example 4.8. Find the height of a golden pyramid if its base has sides 2 inches long.

Figure 4.4: A golden pyramid with the $(1, \sqrt{\phi}, \phi)$ triangle.



Solution: If a side of the base is 2 inches (in) long, then half the base is 1 inch long, and therefore the height of a triangular face is ϕ inches long. This height is the hypotenuse of the right triangle formed by the height of the pyramid (x) and half of its base (1 inch). Therefore, by the Pythagorean theorem,

$$x^2 = \phi^2 - 1^2 = \phi^2 - 1.$$

But according to the defining equation for ϕ ((4.1)), we have $\phi^2 - 1 = \phi$. Therefore,

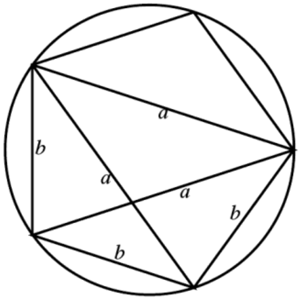
$$x = \pm\sqrt{\phi},$$

and because the height should be positive, we obtain $x = \sqrt{\phi} \simeq 1.272$ in. □

Practice problem 4.9. What is the ratio of a diagonal of a regular pentagon to its side? *Answer:* ϕ.
Hint: Take any four of the five vertices of the regular pentagon and apply *Ptolemy's theorem*, which says: for any quadrilateral whose vertices lie on a circle, the product of the two diagonals equals the sum of the products of the opposite sides.

The first mathematical definition and study of the golden ratio can be found in *Elements* by Euclid (c. 300 BC). However, the work that made

Figure 4.5: A regular pentagon inscribed in a circle.



ϕ famous by emphasizing its aesthetic and even "divine" properties was *De Divina Proportione* by the Italian mathematician and Franciscan friar Luca Pacioli. Published in Venice in 1509 and illustrated by Leonardo da Vinci, the book had a considerable influence on the aesthetics of the Renaissance.

Although multiple references to the presence of the golden proportion in nature and art exist, they should be taken with caution. Computing the ratio of dimensions of an object includes measuring the dimensions with a certain precision and then performing the division. If the rational number thus obtained is more or less close to ϕ, what error should be admissible to claim that it is in fact the golden ratio? For example, it is often claimed that a US drivers license or a credit card is designed as a golden rectangle. If you measure the dimensions, you get an aspect ratio of approximately 1.585, which might or might not be an approximation of ϕ. Similar overeagerness sometimes occurs in the analysis of works of art and architecture. However, *not all that glitters is gold.*

Example 4.10. Does the right triangle with sides $3, 4$, and 5 units have anything to do with the golden ratio?

Solution: Consider the ratios of the sides: $\frac{4}{3} \simeq 1.33333$, $\frac{5}{4} = 1.25$, $\frac{5}{3} \simeq 1.66667$. The last ratio is within 3% of ϕ. Of course, this is not surprising because 3 and 5 are two consecutive Fibonacci numbers. However, the use of a triangle with ratios close to $(3, 4, 5)$ in architecture might originate in its simple rational proportions rather than its relation with ϕ. $\square$

The golden ratio in plants

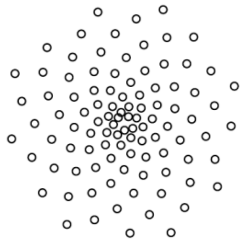
It is universally accepted that the golden ratio seems to be pleasing to the human eye. The reasons for this must be found in nature, and indeed here is one of the manifestations of ϕ in the real world that can be mathematically justified. Namely, the number ϕ appears to lie in the foundation of *phyllotaxis*, the arrangement of leaves, seeds, or florets in many plants.

A visible manifestation of the golden ratio in phyllotaxis is in the number of the clockwise and counterclockwise (right and left) spirals apparent in the structure of composite flowers, leaf arrangements, and seed heads. In about 85% of plants, these numbers are usually two consecutive Fibonacci numbers: 5 and 8, 8 and 13, and so on. Most remarkably, this fact is independent of the biological species (within the 85%; the remaining 15% of plants use entirely different structural arrangements), manifested as well in a sunflower, as in a pineapple or a pine cone. Which pair of Fibonacci numbers appears in a particular case depends on the relative size of the seed with respect to the size of the seed head.

The same pattern can be obtained by numerical simulation, assuming that the angle between the directions from the stem to any two consecutively sprouting seeds (the *divergence angle*) is always the same and corresponds to the splitting of the complete revolution in the golden ratio 1 to ϕ. This *golden angle* equals $\frac{1}{1+\phi} \cdot 2\pi$, or approximately 137.5078 degrees. As we saw earlier in this chapter, the ratios of consecutive Fibonacci numbers converge to the golden ratio. This implies that a Fibonacci number of seeds arranged at the golden ratio divergence angle will wrap almost precisely in several *complete* revolutions around the center, and thus provide starting points for the spirals. For example, $\frac{1}{1+\phi}$ is approximated by $\frac{1}{1+\frac{8}{5}} = \frac{5}{13}$. Therefore 13 seeds get arranged almost precisely in 5 complete revolutions; the next 13 seeds will get pretty close to these; and so on, forming the 13 spirals. The next best approximation of $\frac{1}{1+\phi}$ is $\frac{8}{21}$, so the 21 seeds get wrapped around almost precisely in 8 revolutions, and can serve as starting points for another family of spirals. The two families of spirals are turning opposite ways because $\frac{5}{13}$ is greater, and $\frac{8}{21}$ smaller, than $\frac{1}{1+\phi}$ (if the divergence angle was a fraction of 2π, we would get rays instead of spirals). In fact, any Fibonacci number of seeds gives rise to a family of spirals, but only two families are easy to discern, depending on the size of the seeds with respect to the size of the seed head.

Thus it is apparent that most plants use the golden ratio to determine their growth algorithm. The most prominent theory explaining this phenomenon is quite sophisticated. It considers a plant as a growing system

Figure 4.6: Numerical simulation with the golden angle: 13 right and 21 left spirals.



where the seeds, leaves, or florets are mutually repelling: the second seed appears as far as possible from the first seed and the stem, the third as far as possible from the two existing seeds and the stem, and so on. The development of such a system is determined by the postulate requiring it to stabilize into a state with minimal energy. It can be shown, both theoretically and experimentally, that such a model achieves its stable state as the divergence angle tends to $\frac{1}{1+\phi} \cdot 2\pi$. An interested reader can find an exposition of this theory in more advanced texts.[1]

We choose here to discuss another argument for the prevalence of the golden ratio phyllotaxis, which is derived from the mathematical properties of ϕ as an irrational number. Plants that follow a simple rule to best use available space for structural elements have an evolutionary advantage, thus assuring the prevalence of a particular phyllotaxis in nature. Therefore, a plant with multiple structural elements of similar size and nature (for example, a pinecone comprising many relatively small seeds) would benefit from a simple and universal algorithm that would allow it to grow indefinitely in size while ensuring the best packing of its structural elements, and their equal exposure to sunlight, dew, and rainwater.

For example, let us consider a pinecone. The seeds originate at the core axis of the pinecone, and grow away from it as the cone grows. To minimize the complexity of the process, let us assume that the pinecone adopts the following rules:

1. The divergence angle between each two consecutively sprouted seeds

[1] For example, see M. Livio, *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number,* Broadway Books, 2002.

Figure 4.8: Sunflower: 34 left and 55 right spirals (photo by Chris Darling).

rational approximations is the *universal* solution to the problem (Figure 4.7 right, ratio 1 to ϕ): it guarantees the optimal arrangement according to the rules (1)–(4) across the different sizes of plants, their seeds, leafs, or florets.

The appearance of two consecutive Fibonacci numbers as the numbers of left and right spirals in plants confirms that the angle of divergence constitutes the fraction of a complete circle corresponding to the $1 : \phi$ ratio. This particular seed, leaf, or floret arrangement in plants is called the *Fibonacci phyllotaxis*. Figure 4.7 shows the advantages of the Fibonacci phyllotaxis by numerical simulation: when the golden angle is used for divergence, the packing of the seeds is the best, and alignment is sufficiently random for good access to light. The 13 right and 21 left spirals are prominent in the rightmost pattern.

Counting carefully, you can find exactly 34 left and 55 right spirals in the photo of the sunflower in Figure 4.8.

Practice problem 4.11. Show that ϕ equals the *infinite continued fraction*:

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\dots}}}}.$$

Solution: Denote the continued fraction (i.e., the second term on the right-hand side) by x and notice that $1 + \frac{1}{x} = x$. Multiplying by x, we get

$x + 1 = x^2$, the same equation that defines the golden ratio. Because the continued fraction is clearly a positive number, it follows that $x = \phi$.

Practice problem 4.12.

1. Check that ϕ satisfies the recursive relation

$$\phi^n = \phi^{n-1} + \phi^{n-2}$$

   for any positive integer $n \geq 2$.

   Solution: Indeed, if we multiply (4.1) by ϕ, we get $\phi^3 = \phi^2 + \phi$. Multiplying by ϕ again gives $\phi^4 = \phi^3 + \phi^2$, and so on.

2. Using the recurrence above, we can express ϕ^3 in terms of the first power of ϕ and integers. We know that $\phi^2 = 1 \cdot \phi + 1$ by (4.1), so:

$$\phi^3 = \phi^2 + \phi = (\phi + 1) + \phi = 2\phi + 1.$$

   Similarly, check that $\phi^4 = 3\phi + 2$. Now do the same for ϕ^6. (*Answer:* $\phi^6 = 8\phi + 5$.) In all these cases, note that the coefficients (1,1), (2,1), (3,2), (8,5) are two consecutive Fibonacci numbers.

EXERCISES

Question 4.1. Which of the following numbers are irrational?

1. $\dfrac{3\sqrt{2} - 2}{5}$.

2. $\dfrac{3\sqrt{\frac{9}{4}} - 2}{5}$.

Question 4.2. Which of the following numbers are irrational?

1. $\sqrt{27} - \sqrt[3]{27}$.

2. $\sqrt{2 + \sqrt{2 + \sqrt{2 + \dots}}}$.

Question 4.3. Let x and y be two distinct irrational numbers. Can their product $x \cdot y$ be rational? If yes, give an example. If no, explain why.

Question 4.4. A quadratic equation $ax^2 + bx + c = 0$ with a, b, c rational numbers, has only one real solution. Can this solution be irrational?

Question 4.5.

1. What is the length of the $B2$ size?

2. What is the area of the $B3$ size?

3. What is the reduction factor between the lengths of the $A4$ size and the $B6$ size?

Question 4.6. Find the aspect ratio of an imaginary series of paper sizes, say, $X0$, $X1$, $X2$,... such that

1. three portrait-oriented sheets of the $X(n+1)$ size, arranged side by side next to each other, exactly cover two landscape-oriented sheets of the Xn size, arranged one atop the other,

2. for each size, the aspect ratio is the same.

Question 4.7. Suppose we define, based on the An series of international standard paper sizes, a new series of $B'n$ paper sizes by the following conditions:

(1b′) The length of the $B'n$ size is the *arithmetic mean* of the lengths of the $A(n-1)$ and An sizes for all $n = 1, 2, \ldots, 10$:

$$l(B'n) = \frac{l(A(n-1)) + l(An)}{2}.$$

The length of the $B'0$ size is $\frac{\sqrt{2}+1}{2} \cdot l(A0)$.

(2b′) The aspect ratio for all $B'n$ sizes, $n = 0, 1, \ldots 10$, is $\sqrt{2}$.

1. Check that the new $B'n$ paper sizes satisfy Lichtenberg conditions (1) and (2).

2. Is the new paper size $B'4$ larger or smaller than the international standard $B4$ size? *Hint:* express the length of the $B'4$ and $B4$ paper sizes in terms of the length of the $A4$ size.

3. For two positive numbers a and b, which is bigger: $\frac{a+b}{2}$ or $\sqrt{ab}$? Can they be equal? Explain your answer.

Question 4.8. Suppose an interval is divided into two parts, the larger part a units long and the smaller part b units long. In the definition of the golden ratio, we equate the ratio of the larger part to the smaller part with the ratio of the whole to the larger part:

$$\frac{a}{b} = \frac{a+b}{a} \quad \Longrightarrow \quad \frac{a}{b} = \phi.$$

Let us tweak this definition to see what kind of answers we can obtain.

1. Suppose that the ratio $\frac{a}{b}$ equals three halves of the ratio of the whole $(a+b)$ to the larger part a:

   $$\frac{a}{b} = \frac{3}{2} \cdot \frac{a+b}{a}.$$

   Find the ratio $\frac{a}{b}$.

2. Now suppose that the ratio $\frac{a}{b}$ equals five-sixths of the ratio of the whole $(a+b)$ to the larger part a:

   $$\frac{a}{b} = \frac{5}{6} \cdot \frac{a+b}{a}.$$

   Find the ratio $\frac{a}{b}$.

3. Can you guess how the answer for $\frac{a}{b}$ depends on the positive coefficient p in the equation

   $$\frac{a}{b} = p \cdot \frac{a+b}{a} \quad ?$$

   Try to solve this equation for the ratio $\frac{a}{b}$ in terms of p.

4. Give an example of a value of p in the equation above such that the number $\frac{a}{b}$ is rational.

Question 4.9. Show that $\phi = \sqrt{1+\sqrt{1+\sqrt{1+\ldots}}}$. *Hint:* Check that the infinitely embedded square root satisfies equation (4.1).

Question 4.10. Consider the partial fractions $\{c_n\}$ of the continued fraction

$$\phi = 1 + \frac{1}{1+\frac{1}{1+\frac{1}{1+\frac{1}{1+\ldots}}}},$$

where $c_1 = 1$, $c_2 = 1 + \frac{1}{1}$, $c_3 = 1 + \frac{1}{1+1}$, and so on. The numbers $\{c_n\}$ are called the *convergents* of ϕ. Check that they are equal to the ratios of the consecutive Fibonacci numbers, $c_n = \frac{f_{n+1}}{f_n}$. Can you explain why?
Remark: In general, the presentation of an irrational number as a continued fraction gives its best rational approximations, and these are precisely the convergents. In number theory, there is a way to make precise what we mean by "best" here.

Question 4.11. Derive the recursive relation

$$\left(\frac{1}{\phi}\right)^n = \left(\frac{1}{\phi}\right)^{n-2} - \left(\frac{1}{\phi}\right)^{n-1}.$$

Hint: Derive the equation $(\frac{1}{\phi})^2 = 1 - \frac{1}{\phi}$ from the defining equation (4.1) for ϕ, and use it to obtain the recursion.

Question* 4.12. Check that for any $n \geq 2$

$$\phi^n = f_n\phi + f_{n-1},$$

where f_n is the nth Fibonacci number.
Hint: Use the recursive relation in Practice problem 4.12 to derive a presentation for ϕ^n as a sum of ϕ with an integer coefficient and another integer. Then observe that these integers are constructed according to the same recursive relation as the defining relation for the Fibonacci sequence.

Question* 4.13.

1. Using the recursion in Question 4.11, and proceeding as in Practice problem 4.12, obtain the following formula (pay attention to the sign!):

$$\left(-\frac{1}{\phi}\right)^n = f_{n-1} - f_n\left(\frac{1}{\phi}\right).$$

2. From the previous part, derive the following equality:

$$\frac{1}{\phi} = \frac{f_{n-1}}{f_n} - \frac{1}{f_n}\cdot\left(-\frac{1}{\phi}\right)^n.$$

   Can you conclude that for large n, $\frac{f_{n-1}}{f_n}$ approximates $\frac{1}{\phi}$?
   Remark: This is equivalent to the statement we made in the beginning of this chapter: the ratios $\frac{f_{n+1}}{f_n}$ approximate ϕ, and the approximation becomes better as n grows.

Question 4.14. Find a pine cone. Count its left and right spirals. If they are not consecutive Fibonacci numbers, the cone must have experienced a growth disruption. Find another one and repeat!

It is very instructive to take each equality in the above calculation and see why it is true – and whether you need to use one of the above properties of exponents to justify it.

The next example shows that the exponents need not be concrete real numbers – they can be variables as well.

Example 5.2. Simplify $9^b 4^b - 6^{2b}$. Here, b is a fixed (unknown) real number.
Answer: Using the properties of exponents, we compute:

$$9^b \cdot 4^b - 6^{2b} = (9 \cdot 4)^b - 6^{2b} = 36^b - 6^{2b} = (6^2)^b - 6^{2b} = 6^{2b} - 6^{2b} = 0.$$

□

Here is a third example. The general philosophy is to try to simplify as much as you can. Do not be surprised if the overall answer is not as simple an expression as "20" or "0" (as in the above examples).

Example 5.3. Simplify $2^a 5^b / 10^{a+b}$, where a and b are real numbers.
Answer: Using the properties of exponents,

$$\begin{aligned} \frac{2^a \cdot 5^b}{10^{a+b}} = \frac{2^a \cdot 5^b}{10^a \cdot 10^b} = \frac{2^a}{10^a} \cdot \frac{5^b}{10^b} &= (2/10)^a \cdot (5/10)^b = (1/5)^a \cdot (1/2)^b \\ &= (5^{-1})^a \cdot (2^{-1})^b = 2^{-b} 5^{-a}. \end{aligned}$$

□

Example 5.4 (Solving equations involving exponents)**.** Solve the following equations for the unknown real variable:

1. $5(1+x)^3 = 40$.

   Solution: First divide both sides by 5 to get: $(1+x)^3 = 8$. Now take the cube root of both sides. The left-hand side can now be computed using the properties of exponents, to equal:

   $$((1+x)^3)^{1/3} = (1+x)^{3 \cdot 1/3} = 1 + x,$$

   while the right-hand side becomes $8^{1/3} = 2$. So we get: $1 + x = 2$, or $x = 1$.

2. $(3+x)^5 = 32(2x+1)^5$.

Solution: Take the fifth root of both sides – i.e., raise them to the 1/5th power. Using the properties of exponents (do the work!), we can see that

$$(3+x) = (2^5 \cdot (2x+1)^5)^{1/5} = 2 \cdot (2x+1) = 4x+2.$$

This is a linear equation, so moving all terms involving x to the right-hand side, we get:

$$3-2 = 4x - x = 3x.$$

Finally, $x = 1/3$. □

Practice problem 5.5. Simplify each of the following expressions using the properties of exponents above:

1. $12^6/6^{12}$. *Answer:* $1/3^6 = (1/3)^6$.

2. $(25^3 5^t)^{6-t}$ is what power of 5? *Answer:* 5^{36-t^2}.

3. $\sqrt[4]{2}\sqrt{2}\sqrt[4]{8}$. *Answer:* $2^{\frac{1}{4}} \cdot 2^{\frac{1}{2}} \cdot 8^{\frac{1}{4}} = 2^{\frac{1}{4}+\frac{1}{2}+\frac{3}{4}} = 2^{\frac{3}{2}} = 2\sqrt{2}$.

4. $\dfrac{(\sqrt{RS})^7}{R^3S^3}$, where R and S are positive numbers. *Answer:* $\sqrt{RS}$.

5. $\dfrac{\sqrt[6]{27} \cdot \sqrt[6]{16}}{\sqrt{6}\sqrt[6]{2}}$.

   Answer: $\dfrac{\sqrt[6]{27} \cdot \sqrt[6]{16}}{\sqrt{6}\sqrt[6]{2}} = \dfrac{3^{\frac{3}{6}} \cdot 2^{\frac{4}{6}}}{2^{\frac{1}{2}} \cdot 3^{\frac{1}{2}} \cdot 2^{\frac{1}{6}}} = 3^{\frac{1}{2}-\frac{1}{2}} \cdot 2^{\frac{2}{3}-\frac{1}{2}-\frac{1}{6}} = 3^0 \cdot 2^0 = 1.$

Practice problem 5.6. Solve the equations:

1. $6^{3x} = 36^{\frac{5}{4}}$. *Answer:* $x = \frac{5}{6}$.

2. $(2y^2+1)^5 = 243$. *Answer:* $y^2 = 1$, so $y = \pm 1$.

APPLICATION 1

Computations with exponents are necessary in modeling quickly developing processes, such as chain reactions and avalanches.

Example 5.7. High in the mountains, a comparatively small initial impact (a snowball) can cause a larger snowslide and eventually lead to a huge avalanche.

Suppose that during each second, the volume of the sliding snow grows by a factor of 1/3. What is the volume of the snow in the avalanche after 10 seconds? After 30 seconds?

Solution: Denote the initial volume by V_0. Then after 1 second, we have $V_0 + \frac{1}{3}V_0 = \frac{4}{3}V_0$, after 2 seconds $\frac{4}{3}V_0 + \frac{1}{3}(\frac{4}{3}V_0) = \frac{4}{3}V_0(1+\frac{1}{3}) = \frac{4}{3}\cdot\frac{4}{3}\cdot V_0$, and so on. After 10 seconds, the volume is

$$V_{10} = \left(\frac{4}{3}\right)^{10} V_0 \simeq 17.76\, V_0.$$

After 30 seconds, we have

$$V_{30} = \left(\frac{4}{3}\right)^{30} V_0 \simeq 5,600\, V_0.$$

For example, if the initial volume V_0 was 1 cubic foot (7.5 gallons), then after 10 seconds we have 133 gallons, and after 30 seconds, almost 42,000 gallons of sliding snow.

Now suppose we know that the volume of snow in an avalanche grows in such a way that after 30 seconds, it is 100 times (1,000 times) the initial volume. How much does the volume grow per second?

Denote by r the factor of volume growth per second. Then we have an equation:

$$(1+r)^{30} = 100.$$

This implies

$$1 + r = \sqrt[30]{100} \simeq 1.1659,$$

and $r \simeq 0.1659$. In the second case,

$$(1+r)^{30} = 1,000 \quad\Longrightarrow\quad 1 + r = \sqrt[30]{1,000} \simeq 1.2589,$$

and $r \simeq 0.2589$. □

Another avalanche-type system is provided by a rumor spread.

Example 5.8. Suppose a person obtains some valuable trading information. She probably will not share it massively, but she will tip off her five closest friends. The next day, her friends do the same, and so on. How many *more* people will know the information each day during the next week? We calculate:

day 1	1 person
day 2	5 more people
...	...
day 7	$5^6 = 15,625$ more people

At some point in this process saturation will be reached, where all interested people already know the news and others don't care. Saturation is the reason for failure of financial pyramids, which require exponential increases in the number of participants to sustain them. □

Finally, here is an example of a different kind, where proficiency with powers becomes useful.

Example 5.9. Suppose that a store allows three discounts to apply, consecutively:
10% membership discount,
10% seasonal sale discount, and
10% promotional discount.
Another store with the same merchandise offers a flat 28% discount on all sales. Which deal is better?

Solution: Here the main point is that the three discounts in the first store are applied *consecutively* (instead of *simultaneously*). The first discount reduces the price of an item P to $0.9P$, then the second discount reduces it further to $0.9 \cdot 0.9P = (0.9)^2 P$, and the last discount results in a final price of $0.9 \cdot (0.9)^2 P = (0.9)^3 P$. Therefore, a customer would pay $(0.9)^3 P = 0.729P$ in the first store, and only $0.72P$ in the second store. The second deal is better. □

In contrast, if the first store allows the customer to *add* the discounts, then the total discount comes to 30% of the sale price. In this case the first store is clearly offering a better deal than the second.

APPLICATION 2

Raising positive numbers to powers is necessary in finance, in particular in computing *compound interest*. If you take a loan of a thousand dollars from a bank, or deposit a thousand dollars in it, then, after a couple of years, you will either owe or own *more* than that amount. The original amount of money borrowed or deposited is the *principal*, and the extra amount is called the *interest*.

Simple interest

Let us start with *simple interest*, which is, of course, quite simple. If you deposit $\$1,000$ in a bank that has a 5% annual simple interest rate, then the deposited money accumulates interest, \$50 per year (which is 5% of \$1,000). Thus, after one year you have \$1,050, after two years \$1,100, and so on.

Here is the "general formula" for simple interest: if we deposit an initial sum of money A in a bank that offers an annual simple interest rate of $r\%$, then each year the principal accumulates the same amount of interest, $A \cdot r/100$. Thus, the interest that we would have accumulated on A after t years is: $A \cdot r \cdot t/100$. The total amount after t years is given by:

$$A(t) = A\left(1 + \frac{rt}{100}\right).$$

Example 5.10. A few years ago, I deposited \$2,000 in a bank with a simple interest scheme. The money in my account today is \$2,500.

1. If I deposited the money ten years ago, what is the simple interest rate in the bank?

2. If the rate of simple interest is 5%, then how long ago did I deposit the money in the bank?

Solution: In both calculations, we set $A = \$2,000$ and $A(t) = \$2,500$.

1. If the money was deposited ten years ago, then $t = 10$ and r is unknown, so compute:

   $$2,500 = 2,000\left(1 + \frac{r \cdot 10}{100}\right).$$

   Solving for r, $500 = 20 \cdot r \cdot 10$, whence $r = 2.5\%$.

2. In this part, $r = 5\%$ and t is unknown, so:

   $$2,500 = 2,000\left(1 + \frac{t \cdot 5}{100}\right).$$

   Solving for t, $500 = 20 \cdot t \cdot 5$, whence $t = 5$ yr. □

good an investment this was, let us compute the annual compound interest rate that would produce the same growth. We will make a computation for the initial capital investments of \$24 and \$1,000 separately.

When this book was written, the time elapsed since 1626 was $2014 - 1626 = 388$ years. Assuming the annual compounding model, and \$24 initial capital, we have the following equation for the annual interest rate r:

$$24\left(1+\frac{r}{100}\right)^{388} = 8.53 \cdot 10^{10}.$$

Then

$$\left(1+\frac{r}{100}\right)^{388} \simeq 3.5 \cdot 10^{9} \quad \Longrightarrow \quad \left(1+\frac{r}{100}\right) \simeq 1.058 \quad \Longrightarrow \quad r \simeq 5.8\%.$$

If we take \$1,000 as the initial capital, the annual interest rate r is even lower:

$$1{,}000\left(1+\frac{r}{100}\right)^{388} = 8.53 \cdot 10^{10}.$$

Then

$$\left(1+\frac{r}{100}\right)^{388} \simeq 8.53 \cdot 10^{7} \quad \Longrightarrow \quad \left(1+\frac{r}{100}\right) \simeq 1.048 \quad \Longrightarrow \quad r \simeq 4.8\%.$$

In both cases, the interest rate is not much different from actual interest rates offered by banks for long-term investments, and is compatible with a realistic economy growth rate. If the people who sold the island had instead invested in a bank offering an annual interest rate of about 5%, they would have accumulated roughly the same amount of money as the value of Manhattan land today. In this sense, the sale/acquisition of Manhattan seems to be a reasonable deal.

Continuous compounding and the number e

What happens if the number of compounding periods per year is allowed to grow indefinitely? To simplify our computations, suppose that the principal $A = \$1$, $t = 1$, and $r = 100\%$. In this case, the formula reads

$$A(1) = \left(1+\frac{1}{n}\right)^{n}.$$

Let us see how the result depends on n:

$$\begin{array}{ll} n = 1 & A(1) = (1+1)^1 = 2 \\ n = 2 & A(1) = (1+\frac{1}{2})^2 = \frac{9}{4} = 2.25 \\ n = 4 & A(1) = (1+\frac{1}{4})^4 \simeq 2.441 \\ n = 12 & A(1) = (1+\frac{1}{12})^{12} \simeq 2.613 \\ n = 365 & A(1) = (1+\frac{1}{365})^{365} \simeq 2.714 \\ n = 1,000 & A(1) = (1+\frac{1}{1,000})^{1,000} \simeq 2.717 \end{array}$$

It looks like the values of $A(1)$ become closer together as the number of compounding periods grows. In fact, when n grows indefinitely, this sequence *converges*; that is, it becomes as close as we wish to a certain number, namely to the constant e:

$$\left(1+\frac{1}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e = 2.718281828459045\ldots$$

This number plays an important role in mathematics; for example, it appears as the base of the natural logarithm ln, which we will introduce in Chapter 6. One of the first definitions of e was given in the seventeenth century by the Swiss mathematician Jacob Bernoulli. He defined e as the annual growth factor with 100% compound interest rate *compounded continuously.*

If the compound interest rate is $r\%$, the amount accumulated after t years in the case of continuous compounding is

$$A(t) = Ae^{\frac{rt}{100}}.$$

Example 5.13. With the initial capital $A = \$10,000$, how much money will accumulate in $t = 8$ years if the bank compounds continuously with the compound interest rate $r = 3\%$?

Answer: Using the continuous compounding formula, we compute

$$A(8) = Ae^{\frac{3\cdot 8}{100}} = 10,000e^{0.24} \simeq \$12,712.49.$$

□

Strategy for problem solving

In questions involving compound interest, start with the equation for compound interest (monthly, quarterly, n periods per year). Now plug in the numbers that you are given and identify the unknown quantity. Then solve for the unknown quantity, as is done in the examples below. To make your answer meaningful, you should indicate the units of measurement. For instance, if the amount of money is unknown, the answer should be given in a unit of currency (for example, \$). If the interest rate is unknown, the answer should be given as a percentage (%).

When solving the equation, it is a good idea first to isolate the unknown quantity on one side, writing it in terms of the known quantities, and only then use your calculator to compute the answer. This way, even if the numerical answer is wrong, you can get partial credit for the correct expression, and for doing most of the work to solve the problem. If you have to round your answer, then you should do so only at the *end* of your computations, not in the middle. Here is why: compare 1.01^{20} rounded to one decimal place (you get 1.2) with 1.01 rounded to one decimal place *and then* raised to the 20th power (you get $1.0^{20} = 1.0$).

Example 5.14. How much should be deposited today in an account that earns interest at an annual rate of 6%, compounded monthly, so that it will accumulate to \$20,000 in 5 years?

Solution: We have $r = 6$, $n = 12$, $t = 5$, and $A(t) = 20,000$. We need to find A.

$$A(t) = 20,000 = A\left(1 + \frac{6}{1,200}\right)^{12\cdot 5} = A(1 + 0.005)^{60}.$$

Solving for A, we get:

$$A = \frac{20,000}{(1.005)^{60}}.$$

Now use your calculator:

$$A \simeq 14,827.44.$$

Thus, the amount that should be deposited today is \$14,827.44. □

Example 5.15. Suppose I borrow \$10,000 from a bank that compounds quarterly. Determine the annual compound interest rate (to four decimal places) if after one year I owe \$10,824.32.

Solution: Suppose the annual compound interest rate is r%. Using the formula for interest with quarterly compounding,

$$A(t) = 10,824.32 = 10,000\left(1 + \frac{r}{400}\right)^{4\cdot 1} = 10,000\left(1 + \frac{r}{400}\right)^{4}.$$

To solve for r, we need to divide both sides by $10,000$ and then take the fourth root – i.e., raise both sides to the 1/4th power. (Do you see why taking the fourth root is the same as taking the square root of the square root? Use properties of exponents.) Thus,

$$(1.082432)^{1/4} = \left(1 + \frac{r}{400}\right)^{4\cdot(1/4)} = 1 + \frac{r}{400}.$$

In other words, $1 + (r/400) = 1.01999996$. Solving for r, $r = 7.999984 \simeq 8\%$ is the annual compound interest rate (to four decimal places). □

Example 5.16. A bank uses a compound interest model, so that a deposited amount doubles in 18 years. How much should be deposited in the bank today, to obtain \$50,000 in 10 years?

Solution: We need to find the amount of money X such that

$$50,000 = X\left(1 + \frac{r}{n\cdot 100}\right)^{10n} \quad\Longrightarrow\quad X = \frac{50,000}{(1 + \frac{r}{n\cdot 100})^{10n}}.$$

However, neither the interest rate r nor the period of compounding n is given. All we have is the equation:

$$2A = A\left(1 + \frac{r}{n\cdot 100}\right)^{18n}$$

for any deposit amount A. Dividing by A and taking the 18th power root, we derive

$$\left(1 + \frac{r}{n\cdot 100}\right)^{18n} = 2 \quad\Longrightarrow\quad \left(1 + \frac{r}{n\cdot 100}\right)^{n} = 2^{\frac{1}{18}}.$$

This is all we need to solve the problem. Using the properties of exponents, we have:

$$X = \frac{50,000}{(1 + \frac{r}{n\cdot 100})^{10n}} = \frac{50,000}{((1 + \frac{r}{n\cdot 100})^{n})^{10}} = \frac{50,000}{(2^{\frac{1}{18}})^{10}}.$$

Now, $(2^{\frac{1}{18}})^{10} = 2^{\frac{10}{18}} = 2^{\frac{5}{9}}$. Finally,

$$X = \frac{50,000}{2^{\frac{5}{9}}} \simeq 34,019.75.$$

The deposit has to be \$34,019.75. □

Remark: The point to understand here is that the number of years enters the compound interest formula as an exponent. Therefore, if a deposit doubles in 18 years, we can tell right away that in one year, *any* deposit will grow by a factor of $2^{\frac{1}{18}} = \sqrt[18]{2}$. Then the growth factor over 10 years is computed by raising $2^{\frac{1}{18}}$ to the 10th power, to obtain $2^{\frac{10}{18}} = 2^{\frac{5}{9}}$.

Effective annual rate

In dealing with banks, you may sometimes encounter the expression *effective annual rate* of compounding. Suppose I borrow \$10,000 from a bank with an annual rate of 8%, compounded quarterly. After one year, if the money was compounded only annually, then my loan would be \$10,800. But because the compounding is done more frequently, my loan actually increases. See Example 5.15 above for the calculations: the loan is now \$10,824.32 – some twenty-four dollars more!

The effective annual rate essentially measures this discrepancy. It is defined as the simple interest rate that produces the same amount of money at the end of one year, as the stated compound annual rate, compounded n times per year (quarterly, monthly, ...). Thus, if $r\%$ is the compound annual rate, and it is compounded n times per year, then the effective annual rate r_{eff} is computed according to the formula

$$1 + \frac{r_{\text{eff}}}{100} = \left(1 + \frac{r}{n \cdot 100}\right)^n .$$

For instance, in the above example, the loan after one year is \$10,824.32. Hence the effective annual rate is 8.2432%. (Note that the annual rate was given to be 8%.)

Example 5.17. Find the effective annual rate for the compound annual rate of 6% compounded monthly. Round off your answer to three decimal places.
Answer: Using $r = 6$ and $n = 12$, we compute:

$$1 + \frac{r_{\text{eff}}}{100} = \left(1 + \frac{6}{12 \cdot 100}\right)^{12} = (1 + 0.005)^{12} \simeq 1.0616778.$$

Solving this simple equation, we get $r_{\text{eff}} \simeq 6.168\%$. □

In case of the continuous compounding, the effective annual rate is determined by the formula

$$1 + \frac{r_{\text{eff}}}{100} = e^{\frac{r}{100}} .$$

Example 5.18. Let $A = \$5,000$, $r = 4\%$, compounded continuously.

1. How much money will accumulate in $t = 5$ years?

2. Find the effective annual rate.

Question 5.9. Suppose I have \$15,000 to invest, and two banks are offering competing models. Bank A has an annual interest rate of 6.4%, compounded quarterly, while bank B has an annual interest rate of 6%, compounded monthly. Which bank yields greater interest over one year? In other words, which bank has a greater effective annual rate?

Now compute which bank yields greater interest over three years. In fact, the answer remains the same no matter how much money is invested or how many years of compounding one considers. Why is this so? The answer is: compounding monthly or quarterly by a given interest rate is the same as compounding *annually* by its effective rate. Hence to compute the interest for both models, we are basically compounding annually by their effective rates, in both cases. Thus, whichever effective interest rate is greater will *always* yield a greater return.

Chapter 6

Logarithms I: Money grows on trees, but it takes time

To determine the time it takes for an investment to grow to a certain amount according to a compound interest model considered in Chapter 5, we need a new mathematical tool: the logarithm. In mathematics, taking a logarithm is the inverse operation of taking an exponent. In this chapter we will learn to work with the logarithms and discuss their applications, most importantly in finance.

MATH

For positive numbers $a \neq 1$ and B, the *logarithm* $\log_a B$ is defined as the unique solution to the equation:

$$a^x = B \quad \Longleftrightarrow \quad x = \log_a B.$$

By definition, the exponential function a^x and the logarithmic function $\log_a x$ are inverse to each other. This means that if you apply one and then another, you get the same number you started with. More precisely, for any positive numbers $a \neq 1$ and B we have

$$a^{\log_a B} = B,$$

and similarly, for any positive $a \neq 1$ and any real x we have

$$\log_a(a^x) = x.$$

The number a is the *base* of $\log_a$. We will denote

$$\log = \log_{10}, \quad \ln = \log_e,$$

where the constant $e = 2.71828\ldots$ was introduced in Chapter 5.

Example 6.1. We compute: $\log 1,000 = \log_{10} 1,000 = 3$, because in order to obtain $1,000$, you need to raise 10 to the third power. □

Here is a quick reminder of the *laws of logarithms*, algebraic rules of dealing with the logarithms. Let a, B, C be positive numbers, $a \neq 1$, and r a real number. The main properties of the logarithms are the following:

$$\log_a B + \log_a C = \log_a(BC), \qquad \log_a B - \log_a C = \log_a\left(\frac{B}{C}\right)$$

$$\log_a(B^r) = r\log_a B, \qquad \log_a B = \frac{\ln B}{\ln a},$$

$$\log_a 1 = 0, \qquad \log_a a = 1.$$

These properties follow from the laws of exponents listed in Chapter 5 and the definition of logarithms. Below we prove three of the equations.

Example 6.2. Let a, B, C be positive numbers, $a \neq 1$. Show that

$$\log_a B + \log_a C = \log_a(BC).$$

Solution: Let us introduce the notations: $x = \log_a B$, $y = \log_a C$, and $z = \log_a(BC)$. Then we have by the definition of the logarithm:

$$a^x = B, \quad a^y = C, \quad a^z = BC.$$

By the laws of exponents, this implies

$$a^z = a^x \cdot a^y = a^{x+y}.$$

Hence $z = x + y$. Substituting the values, $\log_a B + \log_a C = \log_a(BC)$. □

Example 6.3. Let a and B be positive numbers, $a \neq 1$, and r a real number. Show that

$$\log_a(B^r) = r\log_a B.$$

Solution: Let $x = \log_a B$. Then by the definition of the logarithm we have:

$$B = a^x.$$

Raising both sides to the power r, we get

$$B^r = (a^x)^r.$$

Taking the logarithm base a of both sides, we have

$$\log_a(B^r) = \log_a((a^x)^r) = \log_a a^{xr} = xr = rx.$$

Plugging in $x = \log_a B$, we get

$$\log_a(B^r) = rx = r \log_a B. \qquad \square$$

Example 6.4. The equality $\log_a 1 = 0$ for any positive $a \neq 1$ is equivalent to the property of exponents: $a^0 = 1$, which holds for any positive a. $\square$

All remaining properties can be proven similarly. (Try it!)

EXAMPLES

In what follows, $\log = \log_{10}$, and $\ln = \log_e$. Each of the following questions can be answered using the definition and laws of logarithms.

Example 6.5. Compute or simplify the following expressions:
(a) $\log \frac{1}{100}$, (b) $(\ln 9 + \ln 4)/\ln(6)$.

Solution: (a) This is the power you need to raise 10 to get $\frac{1}{100}$. We have $10^{-2} = \frac{1}{100}$. Therefore, $\log \frac{1}{100} = -2$.
(b) $(\ln 9 + \ln 4)/\ln 6 = \ln(9 \cdot 4)/\ln 6 = \ln(6^2)/\ln 6 = 2$. $\square$

Practice problem 6.6. Compute or simplify the following expressions:

1. $\log 10^{2t+4}$. *Answer:* $2t + 4$.

2. log of 1 million. *Answer:* 6.

3. $\log 1$. *Answer:* 0.

4. $\log 60 - \log 6$. *Answer:* $\log 60 - \log 6 = \log \frac{60}{6} = \log 10 = 1$.

5. $\log(1 + t) + \log(1 - t) - \log(1 - t^2)$. *Answer:* 0.

(a)

$$t = \frac{\ln 1.5}{4 \cdot \ln\left(1 + \frac{6}{4 \cdot 100}\right)} \simeq 6.81 \ \text{ yr.}$$

(b)

$$t = \frac{100}{6} \ln 1.5 \simeq 6.75 \ \text{ yr.}$$

□

Example 6.10. How long does it take for your investment to grow from $A = \$10,000$ to $A(t) = \$12,000$, if the annual interest rate is 4%, compounded quarterly? Give your answer to the nearest quarter of a year.

Solution: We have $r = 4\%$, $n = 4$, $A = \$10,000$, $A(t) = \$12,000$.

$$A(t) = 12,000 = 10,000 \left(1 + \frac{4}{4 \cdot 100}\right)^{4t}.$$

Dividing both sides by 1,000 gives

$$12 = 10(1.01)^{4t} \quad \Longrightarrow \quad \frac{6}{5} = (1.01)^{4t}.$$

Now is a good time to apply the logarithms. You can use the logarithm with any base, because the answer is independent of the base. We choose the natural logarithm because it is a one-touch operation on any calculator:

$$\ln\left(\frac{6}{5}\right) = \ln(1.01)^{4t} = 4t \cdot \ln 1.01 \quad \Longrightarrow \quad t = \frac{\ln 1.2}{4 \ln 1.01}.$$

Finally, use your calculator:

$$t = \frac{\ln 1.2}{4 \ln 1.01} \simeq 4.6 \ \text{ yr.}$$

The time needed for the investment to grow to \$12,000 is more than 4 years and 2 quarters, but less than 4 years and 3 quarters, because $4.6 > 4.5$ and $4.6 < 4.75$. Therefore, to obtain the desired amount, you will have to wait for 4 years and 3 quarters. □

Strength of an encryption key

Here is an example from a completely different domain. An *encryption key*, or a secret password used in computer security, can be thought of as a sequence of zeros and ones of a certain length. For example, a 56-bit key is a sequence of 56 zeros and ones. It can start like this:

$$(0,0,0,1,0,0,1,1,0,1,0,1,0,0,0,0,0,0,0,0,1,0,1,1,0,1,\ldots\ldots\ldots\ldots)$$

How many different 56-bit keys are there? There are 2 choices (0 or 1) for the first entry. For each choice of the first entry, there are 2 choices for the second, for each choice of the first and the second entry there are 2 choices for the third, and so on. If the key is 3 bits long, then we will have $2 \cdot 2 \cdot 2 = 8$ different sequences. You can easily list them:

$$(0,0,0) \quad (0,0,1) \quad (0,1,0) \quad (0,1,1) \quad (1,0,0) \quad (1,0,1) \quad (1,1,0) \quad (1,1,1)$$

Now suppose your encryption key is 56 bits long. Then you have two independent choices for each of the 56 entries. Therefore you have 2^{56} different possible keys.

Now suppose an imaginary intruder – a hacker or an agency – wants to crack the key by brute force, trying all sequences of zeros and ones until the right one is found. Suppose they have a cluster of $1,000$ processors, and each processor can make 10 billion tries per second. How long would it take them to find the key?

With some luck, you don't have to try all 2^{56} combinations before you hit the right one. Let us assume that you have to try half of them. The processor cluster makes $1,000 \cdot 10 \cdot 10^9 = 10^{13}$ tries per second. Let t be the time in seconds required to find the key. Then we have the equation

$$(10^{13}) \cdot t = \frac{1}{2} \cdot 2^{56} \quad \Longrightarrow \quad t = \frac{2^{55}}{10^{13}} = \frac{2^{42}}{5^{13}} \simeq 3,603 \text{ sec.}$$

It would take the intruder only $3,603$ seconds, or about 1 hour, to crack the key. Even if we assume they need to try *all* possible sequences of length 56, it will take them about 2 hours. If we do not want the key to be vulnerable to brute-force attacks of this kind, we can try to make the key longer. How long should it be to ensure that it takes an adversary decades to crack?

Example 6.11. Assume an adversary has a cluster of $1,000$ processors, each capable of checking 10 billion keys per second. How long should the encryption key be to ensure that it takes more than 30 years for the adversary to crack the key by brute force?

Solution: Let k denote the length of the key. Assume as before that the intruder will have to check half of all possible keys before they hit the right one (actually, taking half or 75% won't make much of a difference for the answer). Then they will have to check $\frac{1}{2} \cdot 2^k$ keys. The time we want them to take is

$$t = 30 \text{ yr} = 10,950 \text{ days} = 262,800 \text{ h} = 94,608 \cdot 10^4 \text{ sec}.$$

So if we want to be safe, we will require that it takes them 10^9 seconds. The computer cluster can make $1,000 \cdot 10 \cdot 10^9 = 10^{13}$ tries/sec. We have the following equation for k:

$$10^{13} \cdot 10^9 = \frac{1}{2} \cdot 2^k \quad \Longrightarrow \quad 10^{22} = 2^{k-1}.$$

To solve for k, we need to take a logarithm. For example, we can take $\log_2$:

$$\log_2(10^{22}) = \log_2(2^{k-1}) \quad \Longrightarrow \quad 22 \log_2 10 = k - 1.$$

Then we compute:

$$k = 1 + 22 \log_2 10 \simeq 1 + 73.08 \simeq 74.$$

A 74-bit key would suffice to withstand an attack by this adversary for over 30 years. □

As technology progresses, the processing speed of the electronic devices is likely to increase. We can make the model more realistic if we assume that the adversary's processor speed grows, say, by a factor of 2 every two years (this growth is consistent with an observation of the hardware development known as Moore's law). This assumption will result in a more complicated equation for the length of an encryption key, which will involve a *finite geometric series*. This will be considered in Chapter 9 (see Example 9.6).

We will discuss other mathematical aspects of cryptography and computer security in Chapter 13.

EXERCISES

Question 6.1. In what follows, $\log = \log_{10}$.

1. Simplify, then compute: $\log\left(\frac{10^{11}}{100^{3.5}}\right)$.
2. Compute $\log \sqrt[7]{1,000}$.

3. Solve for x: $10^{(7x+1)} = \sqrt[3]{100}$.

Question 6.2. In what follows, $\log = \log_{10}$.

1. Simplify, then compute: $\log(15^5) - \log(3^3) - \log(5^5)$.
2. Simplify, then compute: $\log 400 - \log 5 - 3\log 2$.
3. Solve for t: $\sqrt{30} \cdot 100^t = \frac{\sqrt{3}}{1{,}000}$.

Question 6.3. Solve the equations:

1. $e^{(2t+1)} = 18$.
2. $(0.5)^{(10x-8)} = 3$.

Question 6.4. Solve the equations:

1. $14^{3x} \cdot 7^{(x-3)} = 9 \cdot 2^{(3-x)}$.
2. $5^{(2x-2)} = 17 \cdot 25^{(-2x)}$.

Question 6.5. In what follows, $\log = \log_{10}$. Solve the equations:

1. $\log(x^2 - 16) - \log(x - 4) = 2$.
2. $\log(x^2 + 5x + 7) = 0$.

Question 6.6. Suppose you borrow \$4,000 from a bank at an annual interest rate of 10%. Suppose the interest accumulated over a certain time (which need not be an integer number of years) is \$500. Find out how long this takes if the bank compounds (a) monthly, (b) quarterly. Give your answer up to the nearest month.

Question 6.7. How long will it take for an investment to triple at an interest rate of 5% compounded
(a) monthly?
(b) continuously (to the nearest tenth of a year)?

Question 6.8. A bank compounds quarterly at an annual interest rate of r%. With this compounding scheme, the amount of \$1,000 grows to \$1,400 in 10 years.

1. Find the compound interest rate r up to two decimal digits of a percent.

2. With the same compounding scheme, how long will it take for a deposit of $\$20,000$ to grow to $\$25,000$? Give your answer to the nearest quarter of a year.

Question 6.9. Suppose a bank compounds continuously at an annual interest rate of 3.5%. How long will it take for an investment to grow by $\$10,000$ if the initial amount is

1. $\$20,000$?

2. $\$30,000$?

3. $\$50,000$?

Give your answers to the nearest month.

Question 6.10. Suppose a bank compounds monthly. How long will it take for a deposit of $\$5,000$ to grow to $\$8,000$ if the annual compound interest rate is

1. 3%?

2. 4.5%?

3. 5%?

Question 6.11. Suppose the adversary has a cluster of processors capable of making 10^{15} tries per second. You consider an encryption key sufficiently safe if it will take the adversary more than 3 years to crack by brute force. How long should the sequence of zeros and ones be to make a sufficiently safe encryption key? Assume that the adversary will have to check approximately half of all possible sequences before hitting the right one.

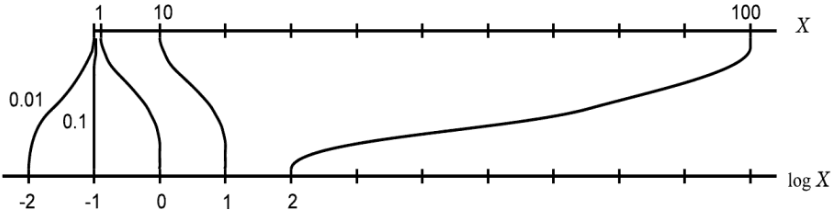
Question 6.12. You have received intelligence that currently the adversary has a cluster of 10,000 processors, each capable of making 5 billion tries per second, and that after 5 years the cluster capability will increase twenty-fold and will not change for the next 5 years. You consider an encryption key sufficiently safe if it would take the adversary more than 10 years to crack by brute force. How long should the sequence of zeros and ones be to make a sufficiently safe encryption key? Assume that the adversary will have to check approximately half of all possible sequences before hitting the right one.

Taking a logarithm makes the difference between large numbers smaller and between small numbers larger. If two positive numbers differ by a factor of 10, their logs (base 10) differ by one unit, whether it is 0.0001 and 0.001, or $100,000$ and $1,000,000$:

$$\log(10B) = \log 10 + \log B = 1 + \log B.$$

As a consequence, the logarithms provide a tool to deal with data of very large range. For example, if the quantities $B = 100$ and $C = 0.01$ need to be marked on the same scale, then visualizing them together on a linear scale is problematic: the difference $B - C = 99.99$ is too large compared to $C = 0.01$. But on a logarithmic scale they become conveniently spaced: instead of C, we mark $\log C = -2$; instead of B, $\log B = 2$. Then the difference between them is only 4 units.

Figure 7.1: Logarithmic versus linear scale.



Tiny quantities become easier to discern, and large quantities can fit into the picture. This makes logarithmic scales convenient in applications with a wide range of data.

Practice problem 7.3. What is the difference between the quantities $A = 1$ and $B = 100,000$ on a logarithmic scale (a) base 10, (b) base 100? *Answer:* (a) $\log(100,000) = 5$, (b) $\log_{100}(100,000) = \frac{\log(100,000)}{\log(100)} = \frac{5}{2}$.

APPLICATIONS

Noise levels

Sounds are created by vibrating objects that generate *sound waves* transporting the energy of the vibration through the medium (for example, air).

In physics, the intensity of a sound is defined as the energy transported by the sound wave past a unit of area per unit of time. It is measured in joules per second per square meter ($\frac{\mathrm{J}}{\mathrm{sec}\cdot\mathrm{m}^2}$), or watts per square meter ($\frac{\mathrm{W}}{\mathrm{m}^2}$). In practice the sounds accessible to detection by the human ear range in intensity from $10^{-12}\frac{\mathrm{W}}{\mathrm{m}^2}$ (the *threshold of hearing*) to $10\frac{\mathrm{W}}{\mathrm{m}^2}$ (pain level). The measurement range of thirteen orders of magnitude suggests the use of a logarithmic scale.

Indeed, the *noise level* of a sound is measured according to the *decibel scale*, which is a logarithmic scale evaluating sound intensity with respect to that of the threshold of hearing (TOH). Given a sound intensity P (in watts per square meter), the decibel (dB) measure of the noise level is given by

$$L_{\mathrm{dB}} = 10\log\left(\frac{P}{P_0}\right),$$

where P_0 is the TOH intensity in watts per square meter.

The decibel scale is consistent with our perception of noise. For example, the sound intensity of a whisper is about 100 times the sound intensity of TOH: $P_{\mathrm{whisper}} = 100P_0$, but we do not think of a whisper as being 100 times louder than the tiniest noise we can hear – it is perceived to be just a little louder. Let us find the decibel level of a whisper:

$$L_{\mathrm{whisper}} = 10\log\left(\frac{P_{\mathrm{whisper}}}{P_0}\right) = 10\log 100 = 10\cdot 2 = 20 \text{ dB}.$$

Here is a table of common sounds together with their estimated decibel levels:

Sound	Noise level (dB)
TOH	0
whisper	20
library	40
normal voice	60
lively street	70
nightclub	100
live rock concert	110
pain level	130

The numbers agree with our intuitive perception of the relative intensity of these sounds.

If we know the decibel level of a sound, we can find its intensity with respect to P_0:

$$L = 10\log\left(\frac{P}{P_0}\right) \quad\Longrightarrow\quad \frac{L}{10} = \log\left(\frac{P}{P_0}\right) \quad\Longrightarrow\quad \frac{P}{P_0} = 10^{L/10}.$$

Example 7.4. Given that the noise level at a nightclub is approximately 100 decibels and that of the normal voice about 60 decibels, find the intensity ratio between the two noise levels.

Solution: Let $L_{\mathrm{n}} = 100$ dB and $L_{\mathrm{v}} = 60$ dB be the decibel noise levels of nightclub music and normal voice, respectively, and P_{n} and P_{v} their kilowatt intensities. Then

$$\frac{P_{\mathrm{n}}}{P_0} = 10^{L_{\mathrm{n}}/10} = 10^{10}; \qquad \frac{P_{\mathrm{v}}}{P_0} = 10^{L_{\mathrm{v}}/10} = 10^{6}.$$

Therefore,

$$\frac{P_{\mathrm{n}}}{P_{\mathrm{v}}} = \frac{10^{10}}{10^{6}} = 10^4 \simeq 10,000.$$

The intensity of a nightclub noise is approximately $10,000$ times that of a normal voice. □

Apparent magnitude of stars

A logarithmic scale of relative *apparent magnitude of stars* has its origins in the work of an ancient Greek astronomer, Hipparchus (190–120 BC). He suggested dividing the visible stars into six levels of magnitude: 1 (the brightest) through 6 (the faintest), and postulated that the increase in one level of magnitude should correspond to a decrease by half in visible brightness. This way of describing the relative brightness of stars was further popularized by Ptolemy (90–168 AD) in his influential treatise *Almagest*. In the nineteenth century, British astronomer Norman Robert Pogson formalized the system by requiring that the stars of magnitude 1 were 100 times as bright as the stars of magnitude 6. An improved and extended version of this scale is used by astronomers today.[1]

Hipparchus' suggestion that the stars of equidistant magnitudes differ in brightness a certain number of times shows that human vision perceives brightness logarithmically. The same assumption lies in the foundation of the Pogson condition – that an increase in magnitude by 5 units should result in a 100-fold decrease in brightness.

Example 7.5. Given that an increase in magnitude by 5 units corresponds to the brightness ratio of $\frac{1}{100}$, find the base of a logarithmic scale suitable to describe the relative apparent magnitude of the stars.

[1] Interested readers can find more details in, for example, *To Measure the Sky: An Introduction to Observational Astronomy*, by Frederick R. Chromey, Cambridge University Press, 2010.

Answer: Let X be the base of the logarithm generating this scale. Then we have:

$$5 = \log_X \frac{1}{100} \quad \Longrightarrow \quad X^5 = \frac{1}{100}$$
$$\Longrightarrow \quad X = \sqrt[5]{\frac{1}{100}} = \frac{1}{\sqrt[5]{100}} = 10^{-\frac{2}{5}} \simeq 0.398.$$

□

Indeed, according to the presently used scale, the apparent magnitude of a star is defined as

$$m = \log_{10^{-\frac{2}{5}}} \left(\frac{F}{F_0}\right),$$

where F is the observed flux (brightness) of the star. The flux is the amount of electromagnetic energy reaching the observer on Earth in unit of time per unit of area, measured in watts per square meter. The quantity F_0 is the flux (brightness) of the star Vega, which is taken as a reference point. The electromagnetic energy comes to the observer in the form of waves of various lengths: X-rays, ultraviolet, visible light, infrared, radio waves, and so on. The ratio of brightness between the sources depends on the wavelengths you take into account: star A can emit more energy in visible light, and star B in the infrared wavelength interval. For the purpose of apparent magnitude, to be consistent with the traditional definition, brightness is understood as the optical broadband flux: the energy emitted across the wavelengths of the visible spectrum.

Example 7.6. Reformulate the definition of the apparent magnitude of stars in terms of log base 10.
Answer: By the base change formula (Example 7.2) we have

$$m = \log_{10^{-\frac{2}{5}}} \left(\frac{F}{F_0}\right) = \frac{\log\left(\frac{F}{F_0}\right)}{\log 10^{-\frac{2}{5}}} = -\frac{5}{2} \log\left(\frac{F}{F_0}\right).$$

□

From now on, we will use the formula for m derived in Example 7.6, which is more convenient for computations.

The scale of apparent magnitude is extended to apply to stars visible only by using a telescope, as well as to certain very bright objects like the planets, the Moon, and the Sun. Because of the negative sign in front of the log, this is an example of a reverse logarithmic scale: the brighter the star, the smaller its magnitude. Very bright objects have negative apparent magnitude. Some well-known celestial bodies and their apparent magnitudes are given in Figure 7.2.

Practice problem 7.7. Compute the apparent magnitude of Sirius, given that its observed flux is 3.87 times that of Vega, and check your answer against the value given in Figure 7.2.
Answer: $m_{\text{Sirius}} = -\frac{5}{2}\log\left(\frac{F_{\text{Sirius}}}{F_0}\right) = -\frac{5}{2}\log 3.87 \simeq -1.47$.

Conversely, given the apparent magnitude of a celestial object, we can find its brightness in the visible spectrum with respect to Vega:

$$m = -\frac{5}{2}\log\left(\frac{F}{F_0}\right) \iff -\frac{2}{5}m = \log\left(\frac{F}{F_0}\right)$$
$$\iff \frac{F}{F_0} = 10^{-\frac{2}{5}m} \simeq 0.398^m.$$

Practice problem 7.8. If the apparent magnitude of the faintest stars visible in an urban neighborhood is $m = 4$, find their relative brightness with respect to Vega. *Answer:* $\frac{F}{F_0} = 10^{-\frac{2}{5}m} = 10^{-\frac{8}{5}} \simeq 0.025$. The brightness of these stars is about 2.5% of the brightness of Vega.

Example 7.9. The mean apparent magnitude of the full Moon is $m_{\text{M}} = -12.74$, and the apparent magnitude of Venus at its maximum brightness is $m_{\text{V}} = -4.89$. What is the ratio of brightness between the Moon and Venus?

Solution: We need to find $\frac{F_{\text{M}}}{F_{\text{V}}}$, where F_{M} is the visible brightness of the Moon, and F_{V} that of Venus. The known apparent magnitudes m_{M} and m_{V} allow us to find the ratios $\frac{F_{\text{M}}}{F_0}$ and $\frac{F_{\text{V}}}{F_0}$:

$$\frac{F_{\text{M}}}{F_0} = 10^{-\frac{2}{5}m_{\text{M}}} = 10^{-\frac{2}{5}\cdot(-12.74)} \simeq 124,738,$$
$$\frac{F_{\text{V}}}{F_0} = 10^{-\frac{2}{5}m_{\text{V}}} = 10^{-\frac{2}{5}\cdot(-4.89)} \simeq 90.$$

Then we can find

$$\frac{F_{\text{M}}}{F_{\text{V}}} = \frac{F_{\text{M}}}{F_0}\Big/\frac{F_{\text{V}}}{F_0} \simeq \frac{124,738}{90} \simeq 1,380.$$

Another solution: We can use the properties of the logarithm to find the required ratio without first computing the brightness of each object with

Answer: By definition of the absolute magnitude, we have

$$
\begin{aligned}
M_{Sun} &= m_{Sun} - 5\log\frac{d}{10} = -26.74 - 5\log(4.85\cdot 10^{-7}) \\
&\simeq -26.74 - 3.43 + 35 = 4.83.
\end{aligned}
$$

If the Sun were placed at the "standard" distance of 10 parsecs from us, it would be just barely visible to the naked eye! □

Even though the distance is factored out from the definition of the absolute magnitude, this quantity describes only the intensity of light emission by the star in the visible spectrum. It does not allow us to estimate other physical parameters of a star, such as size, surface temperature, or total energy emission. Some smaller stars emit more visible light than larger stars.

Earthquake magnitude

A logarithmic scale to measure the relative magnitude of earthquakes was suggested by the American scientist Charles F. Richter in 1934. An earthquake is a vibration of the ground, associated with seismic waves – elastic waves caused by a sudden break or movement of the Earth's crust. It can be characterized by the maximum amplitude of such vibrations, which is measured in units of length, and compared with the reference amplitude. The rough idea behind the *Richter scale* is to calculate the ratio of the amplitude of vibrations A measured at a distance δ from the epicenter, to a reference amplitude function $A_0(\delta)$, and take the logarithm base 10:

$$M_R = \log\left(\frac{A}{A_0(\delta)}\right).$$

The reference amplitude is chosen to be pretty small, so that an earthquake with maximal amplitude of vibrations $A_0(\delta)$ at distance δ from the epicenter is not felt and can be detected only by sensitive instruments. Observations from many seismic stations are combined to determine the position of the epicenter and the distance δ for each station. For convenience, seismographs (instruments used to measure earthquake magnitudes) have the reference amplitude function inbuilt.

In the 1970s the Richter scale was replaced by a more accurate *moment magnitude scale* (MMS), but the principle of the computation remains the same and the results agree whenever both scales are applicable, so that the MMS magnitude is often referred to as the Richter magnitude.

The use of a logarithm base 10 ensures that the increase by one unit in magnitude corresponds to a tenfold increase of the amplitude of an earthquake. Here is a comparative table of earthquake magnitudes.[2]

Magnitude	Effect	Number/year worldwide	Notable earthquake
< 2		1,000,000	
2 to 2.9	Not felt	100,000	
3 to 3.9	Minor	12,000	
4 to 4.9	Light damage	2,000	
5 to 5.9	Moderate damage	200	Long Island, NY, 1884 (5.5)
6 to 6.9	Strong damage; loss of life	20	South Napa, CA, 2014 (6.0)
7 to 7.9	Major; severe damage	3	San Francisco, CA, 1906 (7.9)
≥ 8	Great; total destruction and massive loss of life	< 1	Japan, 2011 (9.0) Chile, 1960 (9.5)

An empirical formula known in geophysics states that the energy of the seismic wave is proportional to the $\frac{3}{2}$ power of the amplitude: $E \sim A^{3/2}$. It follows that an increase by 1 unit in the magnitude corresponds to $10^{3/2} \simeq 31.6$ times the increase of energy released in the earthquake.

Example 7.11. A detonation of about 120 pounds of explosive produces seismic waves of magnitude 2. The magnitude of the 1906 earthquake in San Francisco was 7.9. What is the ratio of the energy released between the San Francisco earthquake and a detonation of 120 pounds of explosive?

Solution: Let A_{SF} and E_{SF} denote the amplitude and energy of the earthquake and $A_{\exp}$ and $E_{\exp}$ of the explosion. Then

$$\log(A_{\mathrm{SF}}) - \log(A_{\exp}) = \log\left(\frac{A_{\mathrm{SF}}}{A_{\exp}}\right) = 7.9 - 2 = 5.9.$$

$$\frac{A_{\mathrm{SF}}}{A_{\exp}} = 10^{5.9} \quad \Longrightarrow \quad \frac{E_{\mathrm{SF}}}{E_{\exp}} = \left(\frac{A_{\mathrm{SF}}}{A_{\exp}}\right)^{3/2} = (10^{5.9})^{3/2} \simeq 7 \cdot 10^{8}.$$

There was approximately 700 million times as much energy released in the earthquake as in the explosion. □

[2] Source: US Geological Survey.

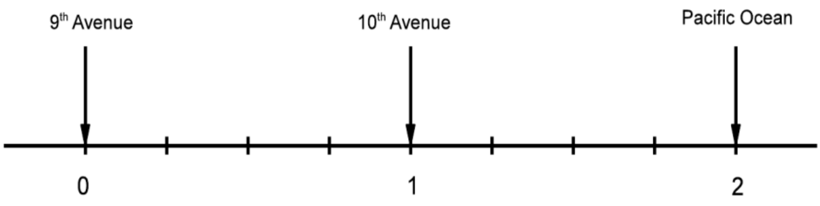
Egocentric scales

In a lighter vein, we can say that logarithms were invented so that we can keep pretending to be the center of the universe. In his famous illustration for *The New Yorker* magazine titled *A View of the World from 9th Avenue* (1976),[3] Saul Steinberg showed a whimsical map of a large part of the western hemisphere – from 9th Avenue in Manhattan all the way to Japan – as seen by a self-absorbed New Yorker. In particular, the distance from 9th to 10th Avenue is about the same as the distance from 10th avenue to the West Coast of the US. To a mathematician, this perception of distances is clearly logarithmic. Let us compute the base of the logarithm a that could have been used to produce such a scale.

Suppose that the distances are measured westward from a certain reference point, and let x denote the distance from this point to 9th Avenue in Manhattan. Taking x to be the reference distance makes 9th Avenue the zero point of our scale:

$$\log_a \frac{x}{x} = \log_a 1 = 0.$$

Figure 7.3: Logarithmic scale of distances as viewed from 9th Avenue.



Let Y and Z denote the distances from 9th Avenue to 10th Avenue and the Pacific Ocean, respectively. These distances are known: $Y \simeq 875\,\text{ft} = 0.166\,\text{mi}$, and $Z \simeq 2,700\,\text{mi}$. Then our logarithmic scale should satisfy the following conditions:

$$\log_a \frac{Y+x}{x} = 1, \qquad \log_a \frac{Z+x}{x} = 2.$$

Therefore,

$$\log_a \frac{Z+x}{x} - \log_a \frac{Y+x}{x} = \log_a \frac{Z+x}{Y+x} = 2 - 1 = 1.$$

[3] You can find the image, for example, at http://www.saulsteinbergfoundation.org/gallery_24_viewofworld.html

From here we find:

$$\frac{Y+x}{x} = \frac{Z+x}{Y+x} = a \qquad \Longrightarrow \qquad \frac{0.166+x}{x} = \frac{2,700+x}{0.166+x} = a.$$

The proportion leads to a quadratic equation for x. In fact, the equation is linear because the x^2 term cancels:

$$0.166^2 + 2 \cdot 0.166 \cdot x + x^2 = 2,700x + x^2$$
$$\Longrightarrow \quad x \simeq \frac{0.0276}{2,699.7} \simeq 0.00001\,\mathrm{mi} \simeq 0.63\,\mathrm{in}.$$

So our self-centered Manhattanite measures all distances from a point 0.63 inches to the east of 9th Avenue (we can assume that this is how close his armchair is to the window). The base of the logarithm is

$$a = \frac{2,700 + 0.00001}{0.166 + 0.00001} \simeq 16,264.$$

To find where any given point is situated according to the 9th Avenue worldview scale ($d_{\text{9th Ave}}$), take the distance d from it to the reference point on 9th Avenue in miles, divide it by 0.00001 miles, and take the logarithm base 16,264.

$$d_{\text{9th Ave}} = \log_{16,264} \frac{d}{0.00001}.$$

Using the base change formula (see Example 7.2), the formula can be rewritten as

$$d_{\text{9th Ave}} = \frac{\log \frac{d}{0.00001}}{\log 16,264} \simeq \frac{5 + \log d}{4.21},$$

where the distance d should be taken in miles.

Example 7.12. How far is Japan from 9th Avenue in New York, according to the 9th Avenue worldview scale?

Solution: The distance from New York to the east coast of Japan is approximately $d = 6,730$ miles. Therefore,

$$d_{\text{9th Ave}}(\text{Japan}) = \frac{5 + \log 6,730}{4.21} \simeq 2.097.$$

Recall that the distance to the West Coast according to this scale is 2. This is why the Pacific Ocean looks like a narrow strip of blue in Saul Steinberg's picture. □

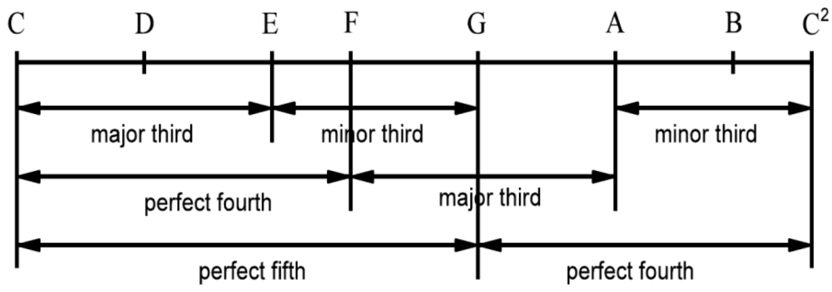
The illustration, initially used as the cover for the March 29, 1976, issue of *The New Yorker*, gave rise to multiple imitations and parodies.

Equal temperament scale in music

In all examples we have encountered so far, logarithmic scales were used mainly for the purpose of accommodating data with a very large range on a single scale. Now we will discuss an application where a logarithmic scale is used primarily for its *multiplicative uniformity*: the property of transforming equal ratios into equal distances.

Traditional European music is based on a seven-pitch octave scale with ratios between the frequencies of the pitches given by ratios of small integers. For example, the octave interval has the ratio of frequencies $2:1$; perfect fifth $3:2$; perfect fourth $4:3$; major third $5:4$; minor third $6:5$. It is believed that our preference for such intervals originates from their presence in nature, which can be explained by the physics of basic natural sounds. The simplest natural drums and whistles that ancient people could have heard (a stick hitting on dry fallen wood, wind blowing in a wooden pipe) produce exactly these kinds of intervals. As early as the third millennium BC, the Babylonian lyre is believed to have had its first seven strings tuned to represent some of the perfect intervals. In classical antiquity there is evidence of the presence of a perfect fourth and major third in the tuning of the strings of the Greek lyre. The heptatonic (seven-pitch) major scale C – D – E – F – G – A – B – C^2 and various versions of heptatonic minor scales gradually gained popularity in European music mainly because of the richness of the harmonic intervals they included.

Figure 7.4: Heptatonic major scale and some of its perfect intervals.

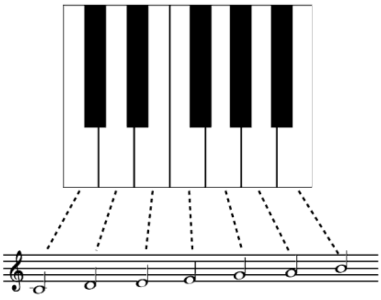


Example 7.13. If the ratio of frequencies between the fourth and the first pitch in the heptatonic major scale is a perfect fourth $\frac{f_F}{f_C} = \frac{4}{3}$, and between the sixth and the fourth – a major third $\frac{f_A}{f_F} = \frac{5}{4}$, then what

The ratio of frequencies between each two consecutive pitches is $2^{\frac{1}{12}}$. Taking the logarithm base 2, we obtain a uniform logarithmic scale with step $\frac{1}{12}$. □

Thus, the octave contains twelve equal intervals, and the ratio between the first and the thirteenth pitch is exactly 2. Other perfect intervals suffer, but not substantially. The heptatonic major scale pitches C-D-E-F-G-A-B are approximated in the twelve-pitch scale by the first, third, fifth, sixth, eighth, tenth, and twelfth pitch, respectively.

Figure 7.5: Twelve-tone equal temperament scale: the white keys correspond to the pitches of the heptatonic major scale.



Example 7.16. Find the ratio of frequencies between the fifth and the first pitch of the twelve-tone equal temperament scale. Which of the perfect intervals does it approximate?

Solution: The ratio of frequencies between the kth and the lth pitch is

$$\frac{f_k}{f_l} = (\sqrt[12]{2})^{k-l}.$$

We compute:

$$\frac{f_{5th}}{f_{1st}} = (\sqrt[12]{2})^4 = 2^{\frac{1}{3}} \simeq 1.2599.$$

The closest perfect interval is the major third $\frac{5}{4} = 1.25$. The error of approximation is relatively small:

$$\frac{1.2599 - 1.25}{1.25} \simeq 0.0079, \quad \text{or} \quad 0.79\%.$$

□