

Advanced Mathematics for Secondary
Teachers:
A Capstone Experience

Curtis Bennett

David Meel

© 2000

August 26, 2001

Contents

1	Introduction	5
1.1	A Description of the Course	5
1.2	What is Mathematics?	6
1.3	Background	9
1.3.1	GCDs and the Fundamental Theorem of Arithmetic . .	10
1.3.2	Abstract Algebra and polynomials	11
1.4	Problems	14
2	Rational and Irrational Numbers	15
2.1	Decimal Representations	18
2.2	Irrationality Proofs	23
2.3	Irrationality of e and π	26
2.4	Problems	30
3	Constructible Numbers	35
3.1	The Number Line	37
3.2	Construction of Products and Sums	39
3.3	Number Fields and Vector Spaces	44
3.4	Impossibility Theorems	54
3.5	Regular n -gons	58
3.6	Problems	60
4	Solving Equations by Radicals	63
4.1	Solving Simple Cubic Equations	66
4.2	The General Cubic Equation	72
4.3	The Complex Plane	74
4.4	Algebraic Numbers	78
4.5	Transcendental Numbers	80

5	Dedekind Cuts	87
5.1	Axioms for the Real Numbers	87
5.2	Dedekind Cuts	91
6	Classical Numbers	105
6.1	The Logarithmic Function	105

Chapter 1

Introduction

1.1 A Description of the Course

This course is intended as a capstone experience for mathematics students planning on becoming high school teachers. This is not to say that this is a course in material from the high school curriculum. Nor is this a course on how to teach topics from the high school curriculum. Rather, this is a course informed by the high school curriculum, by which we mean that the topics in this course are issues raised in the high school curriculum (but rarely dealt with there). Thus students should not expect the course to deal directly with the high school curriculum, although we hope that during the course, students will ask questions that they have concerning that material, including how what we do in the course relates to the curriculum. At the end of many of the sections, we try to include some brief words about how the material from the section might be put to use by a high school teacher, but in truth, mostly this material is here to provide a good background for the teacher, since what is being taught in schools today will probably not be what is taught in 10 years, and almost certainly is not what will be taught in 20 years. Consider that in the 1960s, “New Math” was in vogue with set theory taught at all levels of the curriculum (from first through twelfth grade), clock (or modular) arithmetic taught in elementary school, as well as arithmetic in other bases, while in 1980, the “back to basics” movement took over. The new math was replaced by emphasis on rote skills. Then in 1989, the NCTM standards were introduced, causing changes, such as the elimination of proofs from some high school geometry texts (not what the

NCTM standards called for, but what the textbook writers decided), and now we have the 2000 NCTM standards and their move to include more of the fundamental mathematics, but still with an emphasis on problem solving. Consequently, the goal of the collegiate mathematics education degree is not just to prepare students for teaching now, but to give them the tools to be prepared for teaching twenty years from now.

In a nutshell, this course could be titled “Why do we need all these numbers?” We follow a (mostly) historical development of the real (and Complex) number system, from the Greek Mathematicians through to modern analysis and Dedekind cuts. We begin with a discussion of fractions and rational numbers, and prove that many numbers are irrational. In particular, at the end of chapter 2, we prove that e and π are irrational. Knowing that irrational numbers exist, we then discuss what numbers can be represented as exact lengths using the tools of straightedge and compass. This naturally leads us to prove the classical Greek impossibility theorems on doubling the cube and trisecting the general angle. Given that lengths are not enough, we next move on to whether we can represent all real numbers using radical signs and the standard operations. We show that while we can solve cubic equations this way, these numbers can be deceptive. At the end of that chapter, we give a brief discussion of the impossibility of solving the general quintic equation by radicals, but necessarily, we do not give a proof of this. Finally, having exhausted other methods of defining the real numbers, in the next chapter, we discuss how one defines the real numbers today using Dedekind cuts, and why one is forced to do this. The subsequent chapters are really extras, which we are happy if we get to, but are not necessary to cover in the course if time does not permit.

1.2 What is Mathematics?

The question at the title of this section is extremely difficult. Mathematicians themselves disagree on this question, with some taking a purist view like G.H. Hardy, others taking a more applied approach, and still others giving an “I know it when I see it definition.” In this book, we shall suggest that a brief answer to this question might be that there are four “Ps” to mathematics, pattern, precision, proof, and problem solving.

Mathematics is the science of patterns. The first obvious pattern is that of number. Three people, three hats, and three camels all have something in

common. This is the recognition of a pattern. Seeing how numbers relate to each other usually requires looking for patterns. Of course, patterns become more and more difficult to track down so we come up with more and more complicated techniques to look for them. One of the standard threads of the NCTM standards that fits in here is that of different representations of the same thing. These different representations are often the recognition of the same pattern showing up in two very different items. You have seen this in your mathematics background when you discussed the idea of isomorphism in abstract algebra, congruence in geometry, or even modular equivalence of integers in discrete mathematics. Even elementary school students see this when they first learn that different fractions can stand for the same quantity. One of the greatest discoveries of mathematics is the fundamental theorem of calculus. It is really the recognition that the area under curves and the slopes of curves fit together into a pattern. The idea of calculus is finding the patterns relating these two curves.

To study patterns, you must be precise in your thinking. Thus mathematics is about precision, such as carefully defining what is meant by a term. The English language is fuzzy. By their very nature, words are not precise. As a result, mathematics emphasizes definitions throughout. One often talks about the idea underlying a topic being important, and it is. However, being able to work with this idea is also important, and it is difficult to do so without carefully defined terms.

To verify that the patterns we see actually occur, we turn to proof. There are numerous examples of where people see patterns when they don't really exist. Proof keeps us from treating these patterns as real. For any triangle in the plane, the sum of its angles is 180 degrees. What about on a sphere? If you have studied spherical geometry you know that the sum of the angles of a spherical triangle is greater than 180 degrees, but if you were living on a very large sphere and could only draw relatively small triangles, you might not believe that triangles have angle sum larger than 180 degrees. Why? Because all of the triangles you could easily draw would have angle sum extremely close to 180 degrees, and your measuring tool would not be sufficiently accurate to tell you this. On a more complicated side, every odd number is of the form $4k + 1$ or $4k + 3$ where k is an integer. Here is an interesting question: Is it the case that for any positive integer n the set of primes less than n of the form $4k + 3$ has as many or more members than the set of primes less than n of the form $4k + 1$? If we try and solve this question by example, then we would list the primes of each form and come

up with two lists:

$$4k + 3 : 3, 7, 11, 19, 23, 31, 43, 47, 59, 67, 71, 79, 83, 87, 91, \dots$$

$$4k + 1 : 5, 13, 17, 29, 37, 41, 53, 57, 61, 73, 81, 89, 97, \dots$$

Judging from these lists, the answer is yes for all $n \leq 100$. In fact, the answer is yes for all $n < 10000$. However, the answer is not always yes. For some very large n , the answer is no. Examples teach us a lot, but only with proof can we be sure the generalizations of these examples are true.

The most interesting proofs, however, are those that are wrong. Throughout history, famous mathematicians have published false proofs. Why do they make mistakes? Frequently because some fundamental truth has been misunderstood. Finding a mistake is usually an indication that something hasn't been understood properly. Often, it is finding these mistakes that leads to the most innovative ideas.

Mathematics is also about problem solving. We use the tools of precision and pattern recognition to solve problems. Precision allows us to discover the fundamental knowledge needed to solve a problem, and looking for patterns informs us of how to gain this knowledge. Proofs then provide us with a way to make a convincing argument about why our solution is correct. Indeed, many will argue that this is the most important aspect of mathematics.

Given this definition, why do we teach mathematics? Or rather, what role does mathematics play in a liberal education. Again, even mathematicians disagree in answering this question. The popular answer today is that we teach mathematics because it is everywhere around us, which it is. Mathematics is used in some way in almost all professions. However, this answer might leave us a little uncomfortable because it seems to imply that once computers and machines take over the basic needs, only the programmers would need mathematics education. One answer to this is that many of the uses cannot be easily foreseen, so an experience of discovering the uses will make it easier to find other (sometimes new) uses. A second reason for teaching mathematics is cultural literacy. To make informed decisions, people need to be able to interpret graphs, formulas, and data that they will encounter. For example, how do we understand polling data or assess the risks and benefits of public policy options if we are innumerate? From the other side, an understanding of mathematics allows one to present information quickly and easily. These are important arguments, and it is unfortunate that many

people lack the level of mathematical knowledge to understand information today, even though much of this understanding can be gained with a very basic level of mathematical understanding. Another important reason to study mathematics is that it improves critical thinking skills. It encourages one to think critically in a mathematical way, geometrically, arithmetically, algebraically, and logically. Moreover, proof teaches absolute argumentation so that the validity of other arguments can be weighed against mathematical arguments. G. Polya says it best,

If the (mathematics) student failed to get acquainted with this or that particular geometric fact, he did not miss so much; he may have little use for such facts later in life. But if he failed to get acquainted with geometric proofs, he missed the best and simplest examples of true evidence and he missed the best opportunity to acquire the idea of strict reasoning. Without this idea, he lacks a true standard with which to compare alleged evidence of all sorts aimed at him in modern life.

In short, if general education intends to bestow on the student the ideas of intuitive evidence and logical reasoning, it must reserve a place for geometric proofs. [9]

This view argues that it is not the individual facts of mathematics that matter, but rather the ways of thinking it encourages. The study of mathematics allows us to learn, practice, and master abstract, logical, numerical, and geometric ways of thinking, and to use them to solve problems.

1.3 Background

In this course, we will assume students have familiarity with the concepts from a first course in Abstract Algebra, Linear Algebra, Discrete Mathematics (in particular induction and some basic counting identities), and a full sequence in Calculus. Students may also find it helpful to have had some experience with basic probability.

During the term we will briefly review some concepts from these areas, but students are **strongly encouraged** to have their textbooks from these classes available while reading this text and studying for this class. Moreover, students are responsible to review these areas on their own as necessary.

1.3.1 GCDs and the Fundamental Theorem of Arithmetic

Given integers a and b , we say that a *divides* b if there exists an integer n such that $b = an$, this is written $a|b$. Given the integers a and b , the *greatest common divisor* or gcd of a and b is the largest integer d such that $d|a$ and $d|b$. We now state a very important theorem, which is proven in any discrete mathematics or abstract algebra class.

Theorem 1.1 (GCD is a linear combination theorem) *Let a and b be two non-zero integers. Then there exist integers s and t such that $\gcd(a, b) = as + bt$.*

Recall that an integer $p \geq 2$ is *prime* if it has exactly two positive factors, namely 1 and p itself. Of course, if a is any integer and p is a prime, it follows that $\gcd(a, p)$ is either 1 or p . We can now prove

Theorem 1.2 (Euclid's Lemma) *Let p be a prime and a, b be two integers. If $p|ab$, then $p|a$ or $p|b$.*

Proof: Suppose $ab = pn$ where a, b , and p are as above and n is an integer. Suppose $p \nmid a$. Then $\gcd(a, b) = 1$ and there exist integers s and t such that $1 = as + pt$. Multiplying by b yields $b = abs + ptb = pns + ptb$ and $p|b$ by the distributive law. Thus either $p|a$ or $p|b$.
Q.E.D.

From Euclid's Theorem and induction we obtain the following (which we will not prove):

Theorem 1.3 (Fundamental Theorem of Arithmetic) *Every positive integer greater than 1 has a unique prime factorization. That is, given an integer $n > 1$, then n can be written uniquely as*

$$p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$$

where $p_1 < p_2 < \cdots < p_k$ are primes and m_i is a positive integer for $i = 1, \dots, k$.

1.3.2 Abstract Algebra and polynomials

In this section we will talk about polynomials, roots, extension rings, and fields. In this section, typically we state theorems without proof. Students who wish to look up proofs are encouraged to do so. In general we restrict our attention to polynomial rings and subrings of the complex numbers (and often the real numbers), because these are the cases which will be used throughout the text rather than the more general context discussed in most algebra texts.

A *number field* F is a subset of the complex numbers that is closed under addition, subtraction, multiplication, and (non-zero) division. The finite series

$$p(x) = a_0 + a_1x + \dots + a_nx^n, \quad a_i \in F$$

is said to be a *polynomial* over F . If $a_n \neq 0$, we say that $p(x)$ has degree n . The set of all polynomials over F is denoted by $F[x]$. We have the natural multiplication and addition of polynomials, and $F[x]$ is an integral domain under these operations.

A *root* of the polynomial $f(x)$ is an element $a \in F$ such that $f(a) = 0$.

Theorem 1.4 (Rational Root Theorem) *Suppose $f(x) = a_0 + \dots + a_nx^n$, $a_n \neq 0$ is a polynomial with integer coefficients. If $\frac{b}{c}$ is a root of $f(x)$ with $\gcd(b, c) = 1$, then $b|a_0$ and $c|a_n$.*

Proof: Suppose

$$a_0 + a_1\frac{b}{c} + \dots + a_n\left(\frac{b}{c}\right)^n = 0.$$

Then multiplying the entire equation by c^{n-1} we have

$$a_0c^{n-1} + a_1bc^{n-2} + \dots + a_{n-1}b^{n-1} + \frac{a_nb^n}{c} = 0.$$

Thus the left hand side is an integer. But this implies that a_nb^n/c is an integer. As $\gcd(b, c) = 1$, the Fundamental Theorem of Arithmetic (or a more general form of Euclid's Lemma) implies that $c|a_n$. Multiplying through by $\frac{c^n}{b}$ will allow a similar argument to show that $b|a_0$.

Q.E.D.

The polynomial $p(x)$ is said to *divide* the polynomial $f(x)$ if $f(x) = p(x)q(x)$ for some polynomial $q(x) \in F[x]$, and we write $p(x)|f(x)$. A polynomial $p(x)$ of degree $n > 0$ is *irreducible* if whenever $p(x) = q(x)f(x)$ then

either $\deg(q(x)) = 0$ or $\deg(f(x)) = 0$. Given two polynomials $f(x)$ and $g(x)$, a polynomial $h(x)$ is a *common divisor* of $f(x)$ and $g(x)$ if $h(x)|f(x)$ and $h(x)|g(x)$. The polynomial $h(x)$ is the *greatest common divisor* of $f(x)$ and $g(x)$ if whenever $d(x)$ is also a common divisor of $f(x)$ and $g(x)$ then $\deg(d(x)) \leq \deg(h(x))$. Note that the greatest common divisor of two polynomials is only unique up to multiplication by a constant (when F is a field).

Theorem 1.5 (Division Algorithm for Polynomials) *Let $f(x)$ and $g(x)$ be two non-zero polynomials over some field F . Then there exist unique polynomials $q(x)$ and $r(x)$ in $F[x]$ such that*

$$f(x) = g(x)q(x) + r(x)$$

where either $r(x) = 0$ or $\deg(r(x)) < \deg(g(x))$.

We can now easily see that a is a root of the polynomial $f(x)$ if and only if $f(a) = 0$ by applying the Division Algorithm with $g(x) = x - a$ and plugging in a .

Theorem 1.6 (GCD is a Linear Combination) *Let $f(x)$ and $g(x)$ be in $F[x]$ (where F is a field). Suppose $d(x)$ is the greatest common divisor of $f(x)$ and $g(x)$. Then there exist polynomials $a(x)$ and $b(x)$ in $F[x]$ such that*

$$d(x) = a(x)f(x) + b(x)g(x).$$

This last theorem is used to prove a version of Euclid's Lemma for polynomials, but we state it here in particular because of what it means in the special case where the common divisor is 1.

In particular, if $p(x)$ is an irreducible polynomial of degree n , then the set of *residues* of polynomials upon division by $p(x)$

$$R_{p(x)} = \{a_0 + a_1x + \dots + a_{n-1}x^{n-1} \mid a_i \in F\}$$

is a field under the addition and multiplication operations defined below. Given $f(x)$ and $g(x)$ in $R_{p(x)}$, we define

$$f(x)g(x) = r(x)$$

where $r(x)$ is the remainder of $f(x)g(x)$ divided by $p(x)$. Under this operation, $R_{p(x)}$ is a field which contains the field F as a subfield. The field $R_{p(x)}$

is naturally isomorphic with the quotient field $F[x]/(p(x))$, and $R_{p(x)}$ can be thought of as a set of representatives of the cosets of the ideal $(p(x))$.

A better way of seeing the multiplication in $R_{p(x)}$ is to look specifically at $p(x)$. If $p(x) = b_0 + \dots + b_n x^n$, then in $R_{p(x)}$ we are simply demanding that

$$x^n = -\frac{1}{b_n}(b_0 + \dots + b_{n-1}x^{n-1}).$$

Note that this makes x a “root” of $p(x)$ in the field $R_{p(x)}$.

For example, suppose $F = Q$ is the field of rational numbers and $p(x) = x^2 - 2$. If $p(x)$ is reducible, then $p(x)$ must have a linear factor. But a linear factor would imply that $p(x)$ has a rational root a/b . Using the rational root test, however, we can see that the only possible roots are ± 1 and ± 2 , which clearly are not roots. Hence $p(x)$ is irreducible. Thus, $R_{x^2-2} = \{a + bx | a, b \in Q\}$ is a field. As

$$(ax + b)(cx + d) = acx^2 + (ad + bc)x + bd,$$

in $Q[x]$, and the division algorithm yields

$$acx^2 + (ad + bc)x + bd = (ac)(x^2 - 2) + [(ad + bc)x + bd + 2ac],$$

we have that multiplication in R_{x^2-2} is

$$(ax + b)(cx + d) = (ad + bc)x + bd + 2ac.$$

Next consider the subfield

$$Q[\sqrt{2}] = \{a\sqrt{2} + b | a, b \in Q\}$$

of the real numbers. In this field, multiplication is defined by

$$(a\sqrt{2} + b)(c\sqrt{2} + d) = 2ac + bd + (ad + bc)\sqrt{2}.$$

Comparing this to the multiplication for R_{x^2-2} we see that the two definitions are identical except that we use $\sqrt{2}$ in place of x in $Q[\sqrt{2}]$. In this way, we produce “extension” fields in a natural way.

We point out that $R_{p(x)}$ can be defined for any polynomial $p(x)$, but that it is a field only when $p(x) \in F[x]$ is irreducible over the field F . This is because to get inverses of elements, you need for $p(x)$ to be irreducible over F as you use the GCD is a Linear Combination Theorem.

We also note that if we replace Q with Z (the integers) in the above example, the closure laws still hold. That is the set

$$Z[\sqrt{2}] = \{a + b\sqrt{2} | a, b \in Z\}$$

is closed under multiplication and addition. This is an easy exercise.

1.4 Problems

1. Show $Z[\sqrt{2}]$ is a subring of $Q[\sqrt{2}]$, by showing that multiplication and division are closed operations over $Z[\sqrt{2}]$.
2. Recall that $Q[\sqrt{3}] = \{a + b\sqrt{3} \mid a, b \in Q\}$. Define addition and multiplication in $Q[\sqrt{3}]$ and show how this multiplication is similar to multiplication for R_{x^2-3} , the residue field for $x^2 - 3$ over $Q[x]$.
3. Continuing the previous problem, define a correspondence between $Q[\sqrt{3}]$ and

$$\left\{ \begin{pmatrix} a & 3b \\ b & a \end{pmatrix} \mid a, b \in Q \right\}.$$

Chapter 2

Rational and Irrational Numbers

In this chapter we shall be concerned with understanding the definition of the rational numbers, how this definition can be used to define the mathematical operations, how it corresponds to decimal representations of these numbers, and why we must expand our definition of number beyond the rational numbers. To get us started, take a few minutes and consider the following question.

What do we mean by the number one third?

In thinking about this question, think about the following:

- What role does this number play in different contexts?
- How does our meaning work with the mathematical operations (addition, multiplication, etc.)?
- What makes the concept of fractions difficult for students to understand?

In answering the first of these three questions, you probably discovered that we have many different contexts in which we use fractions. For example, we can think of one-third of thirty objects as being ten, or we can think of one third of a stick, or we simply have a point on the number line. Other answers

might be $\frac{1}{3}$ or $\frac{2}{6}$. The affect of these multiple representations shows up in both the understanding and the proper definition of a rational number.

To effectively carry out the operations using fractions, one third should denote an equivalence class, so that when we add fractions, we choose the most useful element of that class for the problem at hand. When you think about the difficulties many mathematics majors have operating with equivalence classes, it suddenly becomes easier to understand why students who do not like math often have so much trouble adding fractions.

To carefully define the rational numbers, one must use the language of equivalence classes. As usual, we use Z to denote the integers (both positive and negative). Let

$$F = \left\{ \frac{a}{b} \mid a, b \in Z, b \neq 0 \right\}.$$

Let \sim denote the equivalence relation on F given by $\frac{a}{b} \sim \frac{c}{d}$ if and only if $ad = bc$. For the remainder of this section, when we use the word fraction, we shall mean an element of F .

We define the operations of addition and multiplication on F by

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd}, & \text{and} \\ \frac{a}{b} \cdot \frac{c}{d} &= \frac{ac}{bd}. \end{aligned}$$

There is no guarantee that the definitions yield fractions in lowest terms. In fact, under our definition, $\frac{1}{2} + \frac{1}{2} = \frac{4}{4}$, which is then equivalent to $\frac{1}{1}$. This example illustrates how the formula differs from what you “know” is the correct definition for adding fractions with the same denominator, namely $\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$. Our definition does give an equivalent answer $\frac{ac+bc}{c^2}$, since

$$c(ac + bc) = ac^2 + bc^2.$$

Definition. The set Q of rational numbers is the set of equivalence classes of F under the operations (using $\left[\frac{a}{b} \right]$ to denote the equivalence class of $\frac{a}{b}$) $\left[\frac{a}{b} \right] + \left[\frac{c}{d} \right] = \left[\frac{ad+bc}{bd} \right]$ and $\left[\frac{a}{b} \right] \cdot \left[\frac{c}{d} \right] = \left[\frac{ac}{bd} \right]$.

These definitions of addition and multiplication for rationals are forced upon us so that when equivalent fractions are added, we get as our answers equivalent fractions. For example:

$$\begin{aligned} \frac{2}{3} + \frac{4}{5} &= \frac{22}{15} \\ \frac{4}{6} + \frac{12}{15} &= \frac{132}{90}, \end{aligned}$$

and $132 \cdot 15 = 1980 = 22 \cdot 90$. Let's prove this in general for addition.

Proof: Suppose $\frac{a}{b} \sim \frac{a'}{b'}$ and $\frac{c}{d} \sim \frac{c'}{d'}$. Adding, $\frac{a}{b}$ and $\frac{c}{d}$ we get $\frac{ad+bc}{bd}$. Similarly, adding, $\frac{a'}{b'}$ and $\frac{c'}{d'}$ we get $\frac{a'd'+b'c'}{b'd'}$. We now simply need to check that the two answers are equivalent. Multiplying out we get

$$\begin{aligned} (ad + bc)(b'd') &= (ab')dd' + bb'(cd') \\ &= a'bdd' + bb'c'd \quad \text{since } \frac{a}{b} \sim \frac{a'}{b'} \text{ and } \frac{c}{d} \sim \frac{c'}{d'} \\ &= (a'd' + b'c')(bd). \end{aligned}$$

But this implies the answers are equivalent. Q.E.D.

Aside:

Of course, when you learned to add fractions, you didn't think in terms of equivalence classes, but you did spend a long time getting used to the idea that $1/3 = 2/6 = \dots$. So why go through this proof at all? A common problem for precalculus students is how to add **rational functions**, that is fractions of polynomials. Addition in this case appears strange until you see that it only mimics the rational case. What's more, the beautiful thing about this definition for addition is that you avoid the difficulty of finding a least common denominator (l.c.d.). All you really need to do is follow the routine which allows you to get away with any common denominator. This doesn't mean you should teach students **not** to find an l.c.d., but rather you should point out that there are other ways to add fractions. Of course with large numbers and polynomials, least common denominators **do** make the computations easier which is the reason we teach them in the first place.

Now let us think a little about the proof of why this works. It turns out that the proof really depends on two steps that are hidden by the algebra. Thus while the algebra is rather easy to do, the actual "reason" behind doing it is harder to find. This reason, however, is crucial to understanding the process and making it less magical. Most students in this class understand the reason without making it public, but as a teacher, one cannot afford to do this, and having multiple ways of explaining ideas helps enormously. Thus, as silly as it may seem to give two proofs of the same fact that we already agree is correct, let us look at a second proof where we build the proposition up one step at a time.

Proof: First note that for any fraction $\frac{e}{f}$ and non-zero integer x , $\frac{e}{f} = \frac{ex}{fx}$. Hence, if $\frac{a}{b} \sim \frac{a'}{b'}$ and $\frac{c}{d} \sim \frac{c'}{d'}$, then

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd} \\ &= \frac{(ad + bc)b'd'}{bdb'd'} \\ &= \frac{(ab')dd' + (cd')bb'}{bdb'd'} \\ &= \frac{(a'b)dd' + (c'd)bb'}{bdb'd'} \\ &= \frac{(a'd')bd + (b'c')bd}{bdb'd'} \\ &= \frac{a'd' + b'c'}{b'd'} \\ &= \frac{a'}{b'} + \frac{c'}{d'}. \end{aligned}$$

Q.E.D.

This second proof is longer and has a few more steps in it. Thus, at the outset it looks more difficult. However, the key step is more obvious. In this case it is the idea of cancellation of a common term in the numerator and denominator, which then allows a simple algebraic calculation and hides the equivalence class idea altogether.

2.1 Decimal Representations

Rather than beginning this section with a question, we begin with a project.

Program your calculator or a spreadsheet to give an arbitrary number of digits for a fraction $\frac{a}{b}$.

That is, write a program so that given a and b , you can find as many digits of the decimal expansion of $\frac{a}{b}$. In trying to do this, you will want to consider several things

- What does it mean to “bring down a zero” when doing long division?

- What portions of the algorithm repeat themselves?

Such a program will prove useful in many different contexts. It also leads us into the question :

What do we mean by the infinite decimal $\alpha = \overline{.2345}$?

When thinking about this question, think about:

- What number does each digit of the decimal correspond to?
- What rational numbers do you know that α lies between?
- What makes infinite decimals hard to understand?

Decimals are really just an extension of place-value arithmetic in base 10. Thus if the k^{th} digit to the right of the decimal is a_k , then this digit corresponds to the value $a_k \cdot 10^{-k}$. Thus the finite decimal $.2345$ is equal to

$$\frac{2}{10} + \frac{3}{10^2} + \frac{4}{10^3} + \frac{5}{10^4} = \frac{2345}{10000}.$$

If we have k repetitions of this, we have

$$.23452345 \dots 2345 = \sum_{l=1}^k 2345 \cdot 10^{-4l},$$

where there are $4k$ digits in the decimal. Of course, if the decimal is infinite, we will have to resort to an infinite sum leading to limits, which we are not quite ready to do. Consequently, we shall settle for a more finite but less enticing definition, namely that $\overline{.2345}$ is a number α such that

$$\sum_{l=1}^k 2345 \cdot 10^{-4l} \leq \alpha \leq \sum_{l=1}^k 2345 \cdot 10^{-4l} + 10^{-4k}$$

for all integers k . Ever since we were in elementary school, we have been told that such a number exists, however, the existence is not at all obvious.

At this point, we shall give a partial definition of an infinite decimal. This definition will be sufficient for finding the correspondence between rational numbers and decimals. If we desire to move beyond rational numbers in a constructive way, however, we shall have more work to do.

Notationally, we represent the infinite decimal $\alpha = .q_1q_2\dots$ by writing $\alpha = \sum_{i=1}^{\infty} q_i \cdot 10^{-i}$ as we did up above for $.2345$, where we understand that q_i is an integer such that $0 \leq q_i \leq 9$ (where $i \geq 1$). We say that the rational number α is represented by the decimal $q_0 + \sum_{i=1}^{\infty} q_i \cdot 10^{-i}$, if and only if for all positive integers k ,

$$q_0 + \sum_{i=1}^k q_i \cdot 10^{-i} \leq \alpha \leq q_0 + \sum_{i=1}^k q_i \cdot 10^{-i} + 10^{-k}.$$

For example, let us check a well-known fraction such as $\frac{1}{3}$. The definition says that $.\overline{3}$ represents $\frac{1}{3}$ if and only if for all positive integers k ,

$$.33\dots33 \leq \frac{1}{3} \leq .33\dots34,$$

where there are k digits on both the left hand and right hand sides. Multiplying both sides by $3 \cdot 10^k$, this is equivalent to

$$10^k - 1 \leq 10^k \leq 10^k + 2,$$

which is of course correct. Thus, at least in this example, our definition satisfies our intuition.

So how do we discover the infinite decimal given the fraction? This brings us back to our first question of the section. That is, we perform long division. But what does this mean? The integer part of $\frac{a}{b}$ is the quotient in the division algorithm for a and b , and we are left with a remainder r_1 . Now the first digit to the right of the decimal is the quotient of $10 * r_1$ divided by b which is an integer between 0 and 9, as $0 \leq 10 * r_1 < 10 * b$. This also gives us a remainder r_2 and we repeat the process. Thus in equations:

$$\begin{aligned} a &= bq_0 + r_1 \\ 10 * r_1 &= bq_1 + r_2 \\ 10 * r_2 &= bq_2 + r_3 \\ 10 * r_3 &= bq_3 + r_4 \end{aligned}$$

and inductively,

$$10 * r_n = bq_n + r_{n+1} \tag{2.1}$$

This process gives you a decimal expansion

$$\frac{a}{b} = q_0 + \sum_{i=1}^{\infty} q_i \cdot 10^{-i}$$

for $\frac{a}{b}$. We need to see that this expansion satisfies our definition. Consider that $r_n = 10r_{n-1} - q_{n-1}b$, and in general $r_k = 10r_{k-1} - q_{k-1}b$. Plugging in inductively, one obtains

$$\begin{aligned} r_n &= 10r_{n-1} - q_{n-1}b \\ &= 10(10r_{n-2} - q_{n-2}b) - q_{n-1}b \\ &= 10^2r_{n-2} - 10q_{n-2}b - q_{n-1}b \\ &= 10^3r_{n-3} - 10^2q_{n-3}b - 10q_{n-2}b - q_{n-1}b \\ &= \vdots \\ &= 10^{n-1}r_1 - (10^{n-2}q_1b + \dots + 10^2q_{n-3}b + 10q_{n-2}b + q_{n-1}b) \\ &= 10^{n-1}a - 10^{n-1}(q_0 + \sum_{i=1}^{n-1} q_i \cdot 10^{-i})b. \end{aligned}$$

Recalling that the division algorithm tells us that $0 \leq r_n < b$, We divide by $10^{n-1}b$ to obtain the equation

$$0 \leq \frac{a}{b} - (q_0 + \sum_{i=1}^{n-1} q_i \cdot 10^{-i}) \leq \frac{1}{10^{n-1}}.$$

Isolating $\frac{a}{b}$ then yields

$$q_0 + \sum_{i=1}^{n-1} q_i \cdot 10^{-i} \leq \frac{a}{b} \leq q_0 + \sum_{i=1}^{n-1} q_i \cdot 10^{-i} + \frac{1}{10^{n-1}}$$

for all n , which is what we desired.

A question arises: Why did we insist on allowing for equality on the right side when we defined the decimal expansion? This is so that $.\bar{9}$ makes sense, but to understand this cryptic comment we will have to wait until a later chapter. For now, suffice it to say that we want $.\bar{4} + .\bar{5} = .\bar{9}$ to make sense. The understanding of the algorithm also makes clear why rational numbers have repeating decimals. After all, there are only b choices for r_i , so at some point we get a remainder we have had before. Once this happens, however,

it must be the case that we have repetition as you will show in a homework exercise.

Now we have a way to produce a decimal expansion from a fraction, but what about the other way? The traditional way of doing this in the classroom is to take a repeating decimal,

$$x = \sum_{k=0}^{\infty} \left(\sum_{i=1}^n q_i \cdot 10^{-i} \right) \cdot 10^{-k} = \overline{.q_1 q_2 \dots q_n},$$

multiply by 10^n , where n is called the *period* of the repeating decimal, and subtract $\overline{.q_1 q_2 \dots q_n}$ from the output. Let $x = \overline{.q_1 q_2 \dots q_n}$. Then we have:

$$10^n x - x = q_1 \dots q_n = \sum_{i=1}^n q_i \cdot 10^{n-i} = \alpha,$$

and hence $x = \frac{q_1 \dots q_n}{10^n - 1} = \frac{\alpha}{10^n - 1}$, and we see that repeating decimals do correspond to rational numbers. (Note that in the above, $\alpha = q_1 \dots q_n$ denotes the number with digits q_1 through q_n .) This works, **but** it isn't really clear what we mean when we talk about multiplying and subtracting infinite decimals. For example, try this out on $\overline{.9}$ and see what happens. At this point you might start to feel a little queasy about what we just did. Actually, things are even worse as it isn't clear that there is only one rational number equal to a given infinite repeating decimal. Hence we need to be careful and make this conversion process precise.

We begin by showing $\frac{q_1 \dots q_n}{10^n - 1} = \frac{\alpha}{10^n - 1}$ corresponds to the infinite decimal $\overline{.q_1 \dots q_n}$. Using α exclusively now, it suffices to check for all m that

$$\frac{\sum_{t=0}^{m-1} \alpha 10^{tn}}{10^{mn}} \leq \frac{\alpha}{10^n - 1} \leq \frac{\left(\sum_{t=0}^{m-1} \alpha 10^{tn} \right) 10^{mn} + 1}{10^{mn}}.$$

Multiply this inequality by $10^{mn}(10^n - 1)$, to obtain the equivalent inequality:

$$\left(\sum_{t=0}^{m-1} \alpha 10^{tn} \right) (10^n - 1) \leq 10^{mn} \leq \alpha \left(\sum_{t=0}^{m-1} \alpha 10^{tn} \right) (10^n - 1) + 10^n - 1.$$

Noting that $\left(\sum_{t=0}^{m-1} \alpha 10^{tn} \right) (10^n - 1) = 10^{mn-1}$ and dividing through by α , the above is equivalent to

$$10^{mn} - 1 \leq 10^{mn} \leq 10^{mn} - 1 + \frac{10^n - 1}{\alpha}.$$

Clearly the first inequality is correct, and the second inequality holds as $\alpha \leq 10^n - 1$, implying that $\frac{\alpha}{10^n - 1}$ does satisfy our definition.

For uniqueness, note that given two rational numbers, $\frac{a}{b}$ and $\frac{c}{d}$, then they are equivalent to rational numbers with a common denominator $\frac{ad}{bd}$ and $\frac{bc}{bd}$. For some power of 10 we have $10^n > bd$. If the two rational numbers have the same decimal expansion out to the n^{th} decimal, however, then their difference is less than $\frac{1}{10^n}$. As n was chosen so that $\frac{1}{bd} > \frac{1}{10^n}$, however, this is only possible if $ad = bc$ and the two rational numbers are equivalent.

Of course, our algorithm gives a unique decimal expansion, but we **have not shown** that every rational number has a unique decimal expansion. A good thing since this is false.

2.2 Irrationality Proofs

The story is that the Pythagoreans believed that all numbers were rational. Then when one of them proved the existence of an irrational number, the Pythagoreans made a sacrifice to the gods. At least that is one story. Another says that the Pythagoreans threw the offending mathematician overboard. In any case, the Greeks certainly knew that not all numbers are rational. The easiest numbers to think about after the rational numbers are square roots of positive numbers. We know that $1^2 = 1$, $2^2 = 4$, $3^2 = 9$, and in general we call an integer m a *perfect square* (or just a square) if $m = a^2$ for some integer a . So, what about the square root of a non-square integer? Is it always rational? Can it ever be rational? As you probably know from your previous classes, the answer is that the square root of a non-square rational number is never a rational number, but why? This is the question we answer in this section (or more correctly, you will answer in the problems to this section).

For m a positive (non-square) integer, the meaning of \sqrt{m} is straightforward. The \sqrt{m} is a number such that $\sqrt{m} \cdot \sqrt{m} = m$. Of course, we have been a little sneaky here. We don't really know what a number is if it isn't rational, but for now we will leave the question of what we mean by real numbers until Chapter 5. Hence the issue for now is to show that no rational number can have its square equal to m . Once we do that, we will know that if these numbers exist, they aren't rational.

The proofs (and reasons) break into two main classes, algebraic proofs, which use prime numbers and prime factorizations, and analytic proofs,

which use arguments based on inequalities. Here we will prove that $\sqrt{2}$ is irrational in 3 different ways. In the homework you will be asked to do other numbers.

We will begin with the algebraic versions:

Proof 1: This is the traditional even-odd proof. If $\sqrt{2} = m/n$ is rational, we can choose integers m and n so that at least one is odd (not divisible by 2). But then we have $n\sqrt{2} = m$, and squaring both sides gives that $m^2 = 2n^2$ and hence m is even (by Euclid's Lemma). Writing $m = 2k$ we have $(2k)^2 = 2n^2$, so that $n^2 = 2k^2$ and hence n is even. But this contradicts our choice of m and n . Hence $\sqrt{2}$ is not rational.

Q.E.D.

One can also prove this by using the prime factorization of 2 or alternatively by using the rational root theorem on the polynomial $x^2 - 2$. Both of these proofs, however, rest on Euclid's Lemma (if p is a prime number, then $p|ab$ implies $p|a$ or $p|b$), so for our purposes, they are really just more complicated versions of the same proof. Of course, when teaching high school, you might want to use one of these other proofs.

The next two proofs are analytic in nature. By this we mean that they have to do with inequalities, and in the case of the third proof, the idea of limits.

Proof 2: If $\sqrt{2}$ is rational, there exists some smallest positive integer q such that $q\sqrt{2} = p$ is an integer. Then

$$\begin{aligned}(p - q)\sqrt{2} &= p\sqrt{2} - \sqrt{2}q \\ &= 2q - p\end{aligned}$$

is also an integer. As $1 < \sqrt{2} < 2$, we have $q < p < 2q$. Hence $0 < p - q < q$ and $p - q$ is positive and smaller than q . But this contradicts the choice of q as the smallest positive integer such that $q\sqrt{2}$ is an integer. Hence a contradiction has been reached and $\sqrt{2}$ is not rational.

Q.E.D.

Proof 3: Suppose $\sqrt{2} = p/q$ with p and q integers. Note that $\sqrt{2} - 1 < 1/2$ since $2 < 9/4$ implies $\sqrt{2} < 3/2$. Hence by choosing n large, we can make $(\sqrt{2} - 1)^n > 0$ as small as we want. However, $(\sqrt{2} - 1)^n = A\sqrt{2} + B$ for some pair of integers A and B as $Z[\sqrt{2}]$ is a ring (see section 1.3.2), and hence is closed under multiplication and addition. But then writing

$$(\sqrt{2} - 1)^n = A\sqrt{2} + B$$

$$\begin{aligned} &= A\frac{p}{q} + B \\ &= \frac{Ap + Bq}{q}, \end{aligned}$$

we have that if it is positive it must be at least $\frac{1}{q}$. This contradicts that $(\sqrt{2} - 1)^n$ can get smaller than $\frac{1}{q}$. Hence $\sqrt{2}$ is not rational. Q.E.D.

Again, one might ask why we need three proofs that $\sqrt{2}$ is irrational. Each illustrates a different property of rational numbers. The first we have already discussed, the second proof ties into the existence of least terms for a fraction again, but this time based on inequality rather than factors, and the third proof shows the fundamental property of the rational numbers concerning how close you can get to zero using just two rational numbers (1 and $\frac{p}{q}$). This last idea turns out to be extremely useful in many number theory proofs. Later, we shall use this idea in the proof of the irrationality of π .

Teaching Aside: At this point, let's think about when we might use these other proofs in teaching high school? The answer to this question depends on a different question. Namely, why do we teach that $\sqrt{2}$ is irrational? There are several different answers to this latter question. One might argue that students need to understand that not every number can be written exactly as a fraction. However, this argument suggests that we need merely tell them that and be done with it rather than give a proof that a specific number is not rational. A better reasoning might be that understanding the proofs of the irrationality of $\sqrt{2}$ helps one to better understand what the properties of fractions and rational numbers are. Using this idea that we prove the irrationality to teach us about properties of numbers, we can then answer the first question as follows: The first proof could be effectively presented when discussing the fundamental theorem of arithmetic to students on prime decomposition, and its derivatives could be presented as applications to the root theorem for polynomials. The second proof is a good way to introduce induction proofs into the high school (something the NCTM standards recommends), and it can help students learn how to multiply equations containing roots, something which is often covered in high school texts when discussing quadratic equations. Finally the third proof would be useful to present in a precalculus class when discussing limits and the completeness

properties of the real numbers.

End of Aside

What we have done here, is to show that if $\sqrt{2}$ is a number, then it is not rational. Of course, we only have that it is a number because we know from our background that it is a real number. While this statement seems obvious to us, the existence of a number like $\sqrt{2}$ might not be clear. For example, what about $\sqrt{-1}$. This is not a real number, so why do we get away with saying that $\sqrt{2}$ is a real number? We will deal more generally with this topic later when we talk about what the real numbers are.

2.3 Irrationality of e and π

We end this chapter by proving that e and π are both irrational. Even more than the previous section, we are going to use our previous knowledge about these two numbers (and calculus too) to show their irrationality. That e was irrational was known and proved by Euler using much the same technique as we use in Theorem ???. The irrationality of π , on the other hand, was first proved by Lambert in 1761 by the use of continued fractions [?]. We will take a different approach using calculus. Rather than using the more standard calculus proof as in [8], we shall follow the approach that Niven takes in [7]. All of these proofs are based on the same ideas as the second and third proofs of the irrationality of $\sqrt{2}$ above. That is, we desire to show that if the number were rational, some sequence must get arbitrarily small and at the same time must be greater than some fixed positive number establishing a contradiction. For the $\sqrt{2}$, we could do this without resorting to the use of calculus. While one can use basic techniques for e (as we do in Theorem 2.1), proving π irrational requires significantly more work, and the calculus approach seems clearest.

Theorem 2.1 *The number e is irrational.*

Proof: We begin by recalling from calculus that

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

By way of contradiction, suppose $e = \frac{a}{b}$ with a and b positive integers. Then $(b!)e$ is an integer. Using the above expression for e , we have

$$b!e = \sum_{n=0}^b \frac{b!}{n!} + \sum_{n=b+1}^{\infty} \frac{b!}{n!}.$$

The first term on the right is an integer since the k^{th} term of the sum, $b(b-1)\dots(k)$ is an integer. Consequently, the second sum is the difference of two integers, and is therefore itself an integer. Moreover, it is clearly positive, so that we have

$$1 \leq \sum_{n=b+1}^{\infty} \frac{b!}{n!}. \quad (2.2)$$

At this point we analyze each term of this sum. If $n \geq b+1$ is an integer, then

$$n! = n(n-1)\dots(b+1)b! \geq (b+1)^{n-b} \cdot b!.$$

Thus

$$\frac{b!}{n!} \leq (b+1)^{b-n},$$

whenever $n > b$ is an integer. Moreover, the inequality is strict if $n > b+1$. Hence equation 2.2 yields

$$\begin{aligned} 1 &\leq \sum_{n=b+1}^{\infty} \frac{b!}{n!} \\ &< \sum_{n=b+1}^{\infty} (b+1)^{b-n} \\ &= \sum_{n=1}^{\infty} (b+1)^{-n}. \end{aligned}$$

This last is a geometric series. Again, from calculus (or some other previous class), the sum of this series is

$$\frac{1}{b+1} \cdot \frac{1}{1 - (b+1)^{-1}} = \frac{1}{b}.$$

Putting this all together, we obtain that $1 < \frac{1}{b}$, a contradiction.

Q.E.D.

The proof for π is more complicated. In order to help the reader gain some familiarity with the type of argument we use for π , we give a second proof that e is irrational. This proof is given in [16].

Theorem 2.2 *The number e is irrational.*

Proof: We begin by establishing for non-negative integer n that there exists non-negative integers A_n and B_n such that

$$I_n = \int_0^1 x^n e^x dx = A_n e + B_n.$$

If $n = 0$, then $I_n = e - 1$, so that $A_1 = 1$ and $B_n = -1$. Working inductively, suppose that $I_n = A_n e + B_n$ where A_n and B_n are integers. Using integration by parts,

$$\begin{aligned} I_{n+1} &= \int_0^1 x^{n+1} e^x dx \\ &= \left[x^{n+1} e^x \right]_0^1 - (n+1) I_n. \end{aligned}$$

Thus $I_{n+1} = e - (n+1)(A_n e + B_n)$. Hence $A_{n+1} = (1 - (n+1)A_n)$ and $B_{n+1} = -(n+1)B_n$, and thus A_{n+1} and B_{n+1} are integers. By induction, for all n there exist integers A_n and B_n such that $0 < I_n = A_n e + B_n$.

Suppose $e = \frac{a}{b}$. If $0 < Ae + B$ with A and B integers, then $Ae + B \geq \frac{1}{b}$. Consequently, $I_n \geq \frac{1}{b}$ for all n . But, we also know that

$$\begin{aligned} I_n &= \int_0^1 x^n e^x dx \\ &\leq \int_0^1 x^n e dx \\ &= \frac{e}{n+1}. \end{aligned}$$

Thus, $\frac{1}{b} \leq I_n \leq \frac{e}{n+1} < \frac{1}{b}$, a contradiction. Thus e is irrational. Q.E.D.

We shall first outline the proof that π is irrational, and then later we shall do the details. Our goal is to find functions $f_n(x)$ so that $\int_0^\pi f_n(x) \sin(x) dx$ can be shown to be an arbitrarily small positive number. Thus our function $f_n(x)$ will play a role similar to the function x^n in the proof that e is irrational. Assuming $\pi = a/b$ is rational, the function we shall choose is

$$f_n(x) = \frac{x^n (a - bx)^n}{n!}.$$

At this point we need several facts about this function.

Lemma 2.3 Suppose $\pi = a/b$. For any non-negative integer k , $f_n^{(k)}(0)$ and $f_n^{(k)}(\pi)$ are both integers, (where $f_n^{(k)}$ denotes the k th derivative of f_n).

Proof: Let k be given, and note by the product rule that $f_n^{(k)}(x)$ is a sum of terms of the form $A_{m,l}x^{n-m}(a-bx)^{n-l}$ where $A_{m,l}$ is an integer multiple of

$$\frac{n(n-1)\dots(n-m+1) \cdot n(n-1)\dots(n-l+1)}{n!}$$

(you should check this!). Thus if a term in the sum for $f_n^{(k)}(0)$ is not 0, it is an integer as in this case $m = n$ (and $l \leq n$). Similarly if a term in the sum for $f_n^{(k)}(\pi)$ is not 0 it is an integer. Consequently, $f_n^{(k)}(0)$ and $f_n^{(k)}(\pi)$ are both integers.

Q.E.D.

The next step in the proof concerns integrating $\int_0^\pi f_n(x) \sin(x) dx$. As you may recall from calculus, to do this requires integration by parts many times. The reader should do this for $n = 1$, $n = 2$, and $n = 3$ with the function above to get a feel for what is going to happen. Once done, read on.

At this point we set

$$F_n(x) = f_n(x) - f_n^{(2)}(x) + f_n^{(4)}(x) + \dots + (-1)^n f_n^{(2n)}(x).$$

By the above Lemma, note that $F_n(0)$ and $F_n(\pi)$ are integers. We compute

$$\begin{aligned} \frac{d}{dx} (F_n'(x) \sin x - F_n(x) \cos x) &= F_n''(x) \sin x + F_n(x) \sin x \\ &= f_n(x) \sin x, \end{aligned}$$

where the last follows as $f_n^{(2n+1)}(x) = 0$ as $f_n(x)$ is a polynomial of degree $2n$. Consequently, by the Fundamental Theorem of Calculus,

$$\int_0^\pi f_n(x) \sin x dx = [F_n'(x) \sin x - F_n(x) \cos x]_0^\pi = F_n(\pi) + F_n(0). \quad (2.3)$$

Thus by the above, this integral is an integer. At this point we state our theorem.

Theorem 2.4 The number π is irrational.

Proof: Suppose $\pi = a/b$ with a and b positive integers. Then for $x \in (0, \pi)$, $f_n(x) \sin x > 0$. Consequently, $\int_0^\pi f_n(x) \sin x dx$ is positive. By Lemma 2.3, this is a positive integer. However, $f_n(x) \leq \frac{a^n \pi^n}{n!}$ for $x \in [0, \pi]$ (check this!), so that $f_n(x) \sin x \leq \frac{a^n \pi^n}{n!}$ in this range. Consequently,

$$1 \leq \int_0^\pi f_n(x) \sin x dx \leq \pi \cdot \frac{a^n \pi^n}{n!}$$

for all n . As a is a fixed positive integer, this is a contradiction. Hence π is not a rational number.

Q.E.D.

These calculus based approaches to the irrationality of π and e will reappear slightly altered when we discuss the transcendence of these numbers. What about teaching the irrationality of π and e . The proofs for each of these requires some knowledge of calculus, and consequently, they cannot be easily presented to high school students in classes below the level of calculus. This doesn't mean that you shouldn't teach the irrationality, rather that you will have trouble getting students to accept the irrationality. This latter will be particularly true for π if your text insists on using $\frac{22}{7} = 3.14285\dots$ for π (one way to combat this is to point out to students that the next best approximations for π are $\frac{333}{106} = 3.141509\dots$ and $\frac{355}{113} = 3.1415929\dots$, and after that you need a denominator with five digits). So as the teacher, you have to sell the need for more mathematics. Rather than saying we don't do these proofs, point out that one of the reasons that calculus holds a central place in science and mathematics is because it allows us to do these proofs relatively easily. After all, historically, π wasn't shown to be irrational until after the invention of calculus.

2.4 Problems

Warm Up Problems

1. Write a program for your calculator (or on a spreadsheet) that takes as input integers a , $b \neq 0$, and $n > 0$ and produces as output the n th digit after the decimal of $\frac{a}{b}$. Use this program to find the 57th digit of $\frac{1}{97}$. What is the period of $\frac{1}{97}$?

2. Using a calculator or spread sheet, find the period of the decimal expansion of $\frac{1}{n}$ for all integers n between 1 and 60. Explicitly find the decimal expansion for $1/19$ and $1/23$.
3. For all n between 1 and 60, find the least positive integer k_n such that n divides $10^{k_n} - 1$ if such an integer exists.
4. We define a *rational function* as a quotient of two polynomials. We then say that $\frac{p(x)}{q(x)}$ is equivalent to $\frac{r(x)}{s(x)}$ if and only if $p(x)s(x) = q(x)r(x)$. Discuss the difference between equivalence and equality in this case.
5. Generalize as many of the three proofs that $\sqrt{2}$ is irrational as you can to prove $\sqrt{7}$ is irrational.
6. Generalize as many of the three proofs that $\sqrt{2}$ is irrational as you can to prove $\sqrt{21}$ is irrational.
7. We showed that the product of any two rational numbers is rational. Use this to show that $\sqrt{8}$ is irrational.
8. Use integration by parts to calculate (in terms of $f_1(x)$, $\sin x$, and $\cos x$) $\int_0^\pi f_1(x) \sin x dx$, where $f_1(x)$ is as in the proof that π is irrational.
9. Use integration by parts to calculate (in terms of $f_2(x)$, $\sin x$, and $\cos x$) $\int_0^\pi f_2(x) \sin x dx$, where $f_2(x)$ is as in the proof that π is irrational.
10. Using that $f_n(x) \leq \frac{a^n \pi^n}{n!}$, show that

$$\int_0^\pi f_n(x) \sin x dx \leq \pi \frac{a^n \pi^n}{n!}.$$

11. Let A_n and B_n be integers such that $(\sqrt{2} - 1)^n = A_n + B_n \sqrt{2}$. Find the pairs (A_2, B_2) , (A_3, B_3) , and (A_4, B_4) . In each case, calculate A_n/B_n .
12. Prove by induction for all $n \in \mathbb{N}$, that $(\sqrt{3} - 1)^n = A_n + B_n \sqrt{3}$ for some integers A_n and B_n .

13. Compare the answers from problems 2 and 3, what do you notice? State a conjecture and prove it. (**Hint:** For the proof of your conjecture, you should think about how we moved from a fraction to its decimal expansion.)
14. Find necessary and sufficient conditions on integers a and $b \neq 0$ such that $\frac{a}{b}$ has terminating decimal expansion. Prove your answer.
15. Using the definition of equivalence of rational functions from problem 4, prove that if $q_1(x)$, $q_2(x)$, $p_1(x)$, and $p_2(x)$ are rational functions with $q_1(x)$ equivalent to $q_2(x)$ and $p_1(x)$ equivalent to $p_2(x)$, then $q_1(x)p_1(x)$ is equivalent to $q_2(x)p_2(x)$.
16. Generalize as many of the three proofs of the irrationality of $\sqrt{2}$ as you can to show that $\sqrt{28}$ is irrational.
17. Suppose a , b , and c are integers such that $\sqrt{a} + \sqrt{b} = \sqrt{c}$. Show that \sqrt{ab} , \sqrt{ac} and \sqrt{bc} are all integers. Using this, show that there exists an integer d such that $\sqrt{a} = a'\sqrt{d}$, $\sqrt{b} = b'\sqrt{d}$, and $\sqrt{c} = c'\sqrt{d}$ with a' , b' , and c' all integers.
18. * Suppose a and n are integers such that $a^2 < n < (a+1)^2$. Prove that \sqrt{n} is irrational.
19. There is a slightly different proof for the irrationality of π given by Ian Stewart, which we outline here. Suppose $\pi = \frac{a}{b}$ with a and b positive integers. Let

$$I_n = \int_{-1}^{+1} (1-x^2)^n \cos(\alpha x) dx.$$

- (a) Use integration by parts to express $\alpha^2 I_n$ in terms of $4n$, I_{n-1} , and I_{n-2} . (Hint: you will need to show that $I_{n-1} - I_{n-2} = \int_{-1}^{+1} (-x^2)(1-x^2)^{n-2} \cos(\alpha x) dx$.)
- (b) Use induction on n to show that

$$\alpha^{2n+1} I_n = n!(P_n \sin(\alpha) + Q_n \cos(\alpha)),$$

where P_n and Q_n are polynomials in α of degree less than $2n+1$ with integer coefficients.

- (c) Set $\alpha = \pi/2$. Letting $J_n = \alpha^{2n+1} I_n / n!$, show that J_n is an integer and that $0 < |J_n| \leq 2a^{2n+1}/n!$.

- (d) Use the above step to establish a contradiction, so that π must be irrational.

Chapter 3

Constructible Numbers

We have now worked through the definition of the rational numbers, the definition of operations on rational numbers, and examples of numbers that we want to exist, that cannot be rational. Let us begin to examine these numbers. First, consider the following question:

What do we mean by $\sqrt{2}$ and how do we know it exists? What about $\sqrt{3}$?

While thinking about this question, you might want to consider a few thoughts.

- Our calculator gives us a decimal, but can it give us a way of knowing all of the digits?
- Is there some other way to describe $\sqrt{2}$?
- How do we know exactly where to place $\sqrt{2}$ on the number line?

You probably came up with one of two usual answers to the question of how do we know $\sqrt{2}$ exists. The first answer might involve looking at a graph of $f(x) = x^2$ and noting that $f(0)$ is less than 2 while $f(2)$ is greater than 2. Since the graph is unbroken, this implies that at some value of x between 0 and 2, we must have $f(x) = 2$. This argument is correct, but to make it precise, we need the *intermediate value theorem* which requires a significantly deeper understanding of what a real number is in terms of decimals. We will get into this answer in chapter 5.

The other answer involves noting that given a square with side length equal to one unit, the diagonal has side length equal to two units by the Pythagorean Theorem. This works for $\sqrt{2}$, but what about $\sqrt{3}$? Once we

have the $\sqrt{2}$, we know that $\sqrt{3}$ units is the length of the hypotenuse of a right triangle having legs measuring 1 unit and $\sqrt{2}$ units. Continuing in this manner, we can then get all the square roots of integers as recognizable lengths.

One of the primary ways of thinking of numbers is that they measure lengths. As seen above, the advantage of this view of numbers is that it allows us to naturally consider at least some irrational numbers. Of course, if it can't let us view rational numbers at the same time, it wouldn't be very helpful. But at least the numbers $\frac{1}{n}$ can be viewed as lengths breaking a unit length into n equal parts, so that $\frac{1}{2}$ can be viewed as the measure of half a unit stick.

The philosopher/mathematician Zeno came up with several paradoxes related to this. Suppose Achilles and a tortoise are having a race, and to make the race fair, Achilles has given the tortoise a head start. When Achilles reaches the point that the tortoise started at, the tortoise will have moved just a little to some new point p_1 . When he reaches the point that the tortoise is at, however, the tortoise will have moved just a bit farther on from where he was. When Achilles reaches the point p_1 , the tortoise will no longer be there, but will be at some new point p_2 , and so on and so on. Thus Achilles will never reach the tortoise.

Another of Zeno's paradoxes is specific to motion: Suppose you shoot an arrow at a target. Before the arrow can get to the target, it must first get halfway. Now before the arrow can get to the target, it must first go half the distance from where it is to the target. Once the arrow has gone $3/4$ of the distance, it must go halfway again, *ad infinitum*. Thus motion is impossible.

These logical paradoxes are very hard to deal with. We know that they cannot be correct, but on the other hand, where is the flaw in the logic. One answer is that length (or time, or whatever) is not infinitely decomposable. Quantum physics offers up the solution that the location of a particle is not precise so that it is meaningless to say that something is exactly halfway to somewhere else. Others will simply point out that motion happens and will then call it a day.

For us these paradoxes actually get at the heart of the question of what we really mean by number and length. Of course, it is fine to define abstract numbers as infinitely decomposable because we don't actually ask for distances. On the other hand, if we do this, what happens to our applications of numbers. Zeno's paradoxes made a great impact upon Greek mathematics. While the Greeks essentially had the idea of limits, they used the process

of *exhaustion* in many proofs, they also separated the concepts of *magnitude* and number. Today, however, we tend to use these concepts interchangeably. In fact, one of the strengths of mathematics, is that it often allows us to look at similar concepts as being essentially the same. We saw this in the last chapter when we had two different ways to look at rational numbers, as ratios and repeating decimals, and we see it again here when we obtain a new notion, that of distance measurement.

3.1 The Number Line

We begin by noting that one can make sense of rational numbers as length in an elegant way. Suppose we have a unit length called 1. The positive integers are then simply the lengths you can obtain by appending copies of your unit length. This is seen naturally in the number line, where we assign to each point on the number line its distance from 0, where if 0 is to the right of the point, the distance is thought of as negative. With this definition, the point at unit distance to the right of 0 is assigned the number 1, the point two units to the right of 0 is assigned 2, and so on.

Given two lengths d_1 and d_2 , we say that they are *commensurable* if there exists a length m with which you can make lengths d_1 and d_2 by repeated copies. In algebraic notation, this means there exists natural numbers n_1 and n_2 such that

$$\begin{aligned}d_1 &= m \cdot n_1, & \text{and} \\d_2 &= m \cdot n_2.\end{aligned}$$

We can now see that if d_1 and d_2 are commensurable then d_2 is a rational multiple of d_1 (as a number), as $d_2 = \frac{n_2}{n_1}d_1$. Hence, we can define the positive rational numbers as the lengths that are commensurable with our given unit length 1. We will refer to a length with which you can make the lengths d_1 and d_2 as a *common yardstick*. Of course, any rational length $\frac{a}{b}$ has a common yardstick with the unit length, namely the length $\frac{1}{b}$.

This understanding of number as lengths also gives us a geometric picture for both the division algorithm and the Euclidean algorithm. In particular, when dividing the number a by b , we are just seeing how many b length yardsticks it takes to make up an a length yardstick. As a result this gives the quotient as the number of complete b length yardsticks we can line up inside of an a length, and the remainder is the bit left over.

The Euclidean algorithm then just takes the remaining yardstick to measure the b yardstick and inductively repeats this process. The beautiful thing about the geometric interpretation is that we no longer need to have integers to perform the Euclidean algorithm.

Let us look at an example using $a = \frac{3}{4}$ and $b = \frac{2}{7}$. The Euclidean algorithm then gives us:

$$\begin{aligned}\frac{3}{4} &= \frac{2}{7} \cdot 2 + \frac{5}{28} \\ \frac{2}{7} &= \frac{5}{28} \cdot 1 + \frac{3}{28} \\ \frac{5}{28} &= \frac{3}{28} \cdot 1 + \frac{2}{28} \\ \frac{3}{28} &= \frac{2}{28} \cdot 1 + \frac{1}{28}.\end{aligned}$$

So that $\frac{1}{28}$ is the largest common yardstick for $\frac{3}{4}$ and $\frac{2}{7}$.

While it seems natural that the largest common yardstick involves the least common denominator, it is worthwhile to note this surprising fact, because of how it came up. The Euclidean algorithm will naturally produce a least common denominator, just as it produces a greatest common divisor, precisely because the two concepts are really just different sides of the same coin. Least common divisors are really just least common multiples of the denominators, and it is always the case that the least common multiple of two numbers multiplied by the greatest common divisor is precisely the product. (Check this using the fundamental theorem of arithmetic.)

In general, if we perform the Euclidean algorithm on two yardsticks of length d_1 and d_2 , it will stop with a length m if and only if m is a common yardstick for d_1 and d_2 . As two steps of the Euclidean algorithm always decrease the yardstick length by at least half (see homework exercise), this tells us that if two lengths are not rational multiples of each other, then we can use them to make as small of a length as we like. Let us collect this in a theorem.

Theorem 3.1 *Let l and k denote the lengths of two line segments. Then there is a common yardstick for these segments if and only if there is a rational number q such that $l = kq$. Moreover, if l and k have no common yardstick, then we can use copies of these segments to make as small a segment as we like. In equations, for all $\delta > 0$ there exist integers m and n such*

that

$$\delta > ml + nk.$$

As a quick example, in the above case, with $l = \frac{3}{4}$ and $k = \frac{2}{7}$, we get $q = \frac{21}{8}$. We have now exhausted what we can do with lengths in one dimension. Can we improve on this using two dimensions?

3.2 Construction of Products and Sums

Let us get back to our basic question. What do we mean by $\sqrt{2}$? Algebraically, we mean a number x that satisfies the equation $x^2 = 2$. But how do we know such a number exists. In fact, how do we know where to put it on the number line? Of course, we can only put it down approximately on the number line, but in theory we can place the square root of two exactly. All we need to do is construct an isosceles right triangle with legs of length 1 unit, and then the Pythagorean theorem tells us that the hypotenuse has length $\sqrt{2}$. This leads us to the idea of *constructible numbers*. For our purposes we shall follow the rules that the ancient Greeks followed. Given a unit length, we say that the positive number a is constructible, if we can construct a segment of length a using just a straightedge (a ruler with no marks) and compass. (We will give a slightly more precise definition much later in the chapter, but for now we shall use this definition.) These rules are arbitrary, and other sets of rules can be developed that also lead to interesting work. For this text, our starting assumptions will be that you know how to copy angles, copy lengths, construct perpendiculars, and bisect lengths and angles using a straightedge and compass. Moreover, we shall assume that the laws of side-side-side (SSS), side-angle-side (SAS), and angle-side-angle (ASA) congruence are all known (and have ideally been thoroughly discussed in a previous course).

Consider the following question:

Given a unit length segment and segments of length a units and b units, can you construct using only a straightedge and compass a segment of length $a \cdot b$ units? What about a segment of length $\frac{a}{b}$ units?

While thinking about this question, consider the following

- What theorems about triangles involve products or quotients?

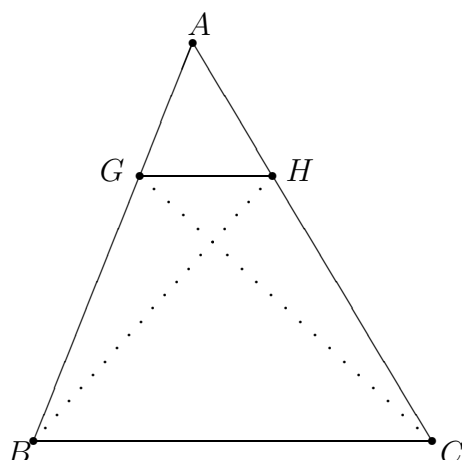
- How can you construct triangles that satisfy these laws?
- How is division really defined?

If you are still having trouble coming up with a way to do this, think about different methods for measuring the height of a flag pole. The traditional mathematical method is to use similar triangles. Using a meter stick, you measure the shadow of the meter stick and the shadow of the flag pole. Then using similar triangles, you know that the height of the flag pole (in meters) is the length of the shadow of the flag pole divided by the length of the meter stick. What we have just done is discovered how to divide two numbers via similar triangles. At this point, you should figure out how to do multiplication by yourself. (This idea of performing multiplication and division of two lengths and getting a length out, seems to have first been done by René Descartes in his book *Geometrie* [15]/)

Since we just used similar triangles, let us go back and review them. Recall that two geometric figures are defined to be *similar* if their corresponding angles are congruent and their corresponding sides are in a common ratio. Conveniently, for triangles it suffices to check that the corresponding angles are congruent. Since all rectangles have four right angles, and yet there are certainly two rectangles that aren't similar, a theorem like the Angle-Angle-Angle (AAA) theorem for similarity is false for polygons with more sides. Why is this theorem true?

Theorem 3.2 (AAA Theorem) *Given triangles ABC and DEF , such that the corresponding angles are congruent, then the triangles are similar.*

Proof: Let triangle ABC be given, r_1 be the ray with end point A through B , and r_2 be the ray with end point A through C . Let G and H be the points on r_1 and r_2 respectively so that $\overline{DE} \equiv \overline{AG}$ and $\overline{DF} \equiv \overline{AH}$.



By side-angle-side theorem for congruence of triangles, triangle AGH is congruent to triangle DEF . (We will accept the SAS theorem as true because we have to start somewhere, and it is an easier theorem to prove.) By the corresponding angles theorem, \overline{GH} is parallel to \overline{BC} . We can now assume that \overline{AG} and \overline{AH} are respectively shorter than \overline{AB} and \overline{AC} . So that triangle AGH is properly contained in triangle ABC . At this point, we need to draw in some auxiliary lines and triangles. In particular, consider triangles GHB and GHC . As they both have base \overline{GH} and height d , the distance between the parallel line segments \overline{GH} and \overline{BC} , the area of GHB is equal to the area of GHC . As triangle AGC breaks into triangles AGH and GHC , and triangle AHB breaks into triangles AGH and GHB , it follows that the area of triangle AGC is equal to the area of triangle AHB . Considering \overline{AH} as the base of triangle AHB and \overline{AC} as the base of triangle ABC , we see that these two triangles have the same height so that the ratio of their areas is $\frac{|\overline{AH}|}{|\overline{AC}|}$. Similarly, the ratio of the areas of triangle AGC and ABC is $\frac{|\overline{AG}|}{|\overline{AB}|}$. Since triangles AGC and AHB have the same area, however, these two ratios are equal. Thus

$$\frac{|\overline{AH}|}{|\overline{AC}|} = \frac{|\overline{AG}|}{|\overline{AB}|}.$$

A similar argument completes the proof.
Q.E.D.

Aside

Why did we put this proof in the book. First of all, why bother with the proof. Probably, we all remember this theorem from an earlier class. Let's think about the following, if you teach this theorem in a high school geometry class. Why do you want students to learn it, why might you want them to see a proof, and why might you need to know a proof? Most students that take high school geometry will not remember this theorem when they are 25, so it is unlikely that most students will need to know this fact when they finish their schooling. Of course, it is probably good for them to have seen this, so if they do need it later, they can relearn it quicker. Maybe this answers the first question. But then, why might you want them to see the proof? Proofs in general are a useful way of thinking. Recall the words of G. Polya, "But if he (the student) failed to get acquainted with geometric proofs, he missed the best and simplest examples of true evidence and he missed the best opportunity to acquire the idea of strict reasoning. Without this idea, he lacks a true standard with which to compare alleged evidence of all sorts aimed at him in modern life. (paragraph) In short, if general education intends to bestow on the student the ideas of intuitive evidence and logical reasoning, it must reserve a place for geometric proofs." Polya's argument is that students don't need to know all geometric facts, but they need to be acquainted with the proof.

Another reason for looking at this particular proof, is that in discussing it, we can illustrate two useful techniques hidden inside of the proof. The main idea of the proof is to use area to talk about length, even though there is no obvious area to consider at the outset. Thus the proof illustrates the possible benefit of changing the problem at hand so that we can apply other ideas to it. This method of problem solving is useful not only in mathematics, but also in solving any difficult problem. The second technique is to consider auxiliary triangles, that is we drew a line that did not appear in the original problem. This technique is extremely helpful in solving geometric problems.

Let's think about algebra now. The whole idea behind the use of variables in algebra is that it allows us to solve many equations at the same time. Hence, a common trick is to solve a problem with simple numbers, and then change some of the numbers to variables. Here we are eliminating information to solve many problems. However, often what one does is to assign variables specific values, solve the specific problems, and then attempt to put the variables back in step by step. In this case, we add in extra information that makes the problem easier to solve.

Leaving the mathematical domain, think about what we do when giving

directions to people. We often add in extra information to make it easier for them to follow the directions. Of course, we call these landmarks.

None of these ideas can come out however, if you don't discuss the proof, either while doing it, or after you have completed the proof. Thus, when teaching something like the AAA proof, it is worthwhile to point out what techniques make the proof possible. In particular, after concluding a proof like this, it will help the students immensely if you go back and help them see what in the proof is worth looking at for review.

End of the Aside

From the AAA-theorem, we can now construct similar triangles, one having sides of length 1 and a , and the other having corresponding sides of length b and x . The length of x must then be $a \cdot b$ giving our product. It is left to the reader to find a technique for the quotient. Of course, constructing sums and products of constructible lengths is easy. So, what does this mean? Before finishing, let us define a negative number x to be constructible if $|x|$ is constructible. Now, we can construct products, sums, differences, and divide by non-zero numbers. Thus:

Theorem 3.3 *The set of constructible numbers forms a field.*

We already know that the set of constructible numbers is larger than the set of rational numbers. Does it cover all the real numbers? In the rational case, we discovered that there were rational numbers that didn't have a rational square root. Let our first goal be to show that every constructible number has a constructible square root.

Theorem 3.4 *If a is a constructible number, then so is \sqrt{a} .*

Proof: We will outline the steps of the construction here and leave the details for the reader. Since we can construct a segment of length a , we can construct a segment of length $1 + a$, which we will denote \overline{AB} , letting C be the point on this segment such that $|\overline{AC}| = 1$ and $|\overline{CB}| = a$. Let O be

the midpoint of the segment $|\overline{AB}|$, which we can construct by bisecting the segment. At this point we can construct the circle centered at O through the

points A and B . Let ℓ be a line perpendicular to \overline{AB} and let D be the point of intersection of ℓ and the circle. By homework 11, angle $\angle ADB$ is a right angle. Then by homework problem 12 $|\overline{DB}| = \sqrt{a}$.
Q.E.D.

3.3 Number Fields and Vector Spaces

Our main goal in this section is to examine whether there are any limits on what numbers we can construct. This study has great historical significance. The ancient Greeks asked the questions of whether you could trisect a general angle with straightedge and compass, whether you could double the cube with straightedge and compass, and whether you could square the circle with straightedge and compass. Let us take these one at a time and see what they mean.

To trisect the general angle means to give an algorithm using a straightedge and compass that will given the input of an angle, produce an angle of measure one third the first angle. This problem seems least clearly associated to what numbers we can construct, but as we shall see later, using angle sum identities for the cosine, we can turn this question into one about constructing lengths.

The proper statement of the doubling the cube question is: Given a segment congruent to the side of a cube, use a straightedge and compass to construct a segment congruent to the side of a cube of twice the volume of the original. Thus, if the first segment has length a , and thus the cube has volume a^3 , we wish to construct a segment of length b so that $b^3 = 2a^3$. Thus, if a is chosen to be 1, b would need to be $\sqrt[3]{2}$. Hence, the question of doubling a cube is equivalent to the question: Given a unit segment, can we construct a segment of length $\sqrt[3]{2}$?

Squaring the circle is shorthand for the problem of given the radius of a circle, can you construct a line segment so that a square having sides congruent to your segment would have the same area as the circle. Given a circle of radius 1, this is equivalent to constructing $\sqrt{\pi}$. As π is a more complicated number than 2, we would expect this to be a significantly harder problem.

While the Greeks could not solve these problems, and indeed, the last of them was only solved in the late 19th century, that isn't to say that they didn't know how to do them with other tools. Indeed, using conic sections

and other curves, they were able to solve these problems (see [?] for references to early proofs for example). They simply could not solve them using the constraints of a straightedge and compass.

Tackling these problems requires a surprising amount of sophistication. The proofs are complicated enough that they are hard to explain to non-mathematicians. Ideally, however, prospective teachers should be able to at least broach the subject with their students, particularly as many students find the existence of impossibility proofs to be fascinating. So, let us begin by reviewing necessary material.

We say that a subset of the complex numbers is a **number field** if it is closed under addition, subtraction, multiplication, non-zero division, and contains at least one non-zero element. Recall that a pair $(V, +)$ is an additive set if V is a non-empty set and $+$ is a binary map from $V \times V$ to V . Given a number field F , an additive set $(V, +)$ and a binary operation $\cdot : F \times V \rightarrow V$, we say that V is a **vector space over F** if

1. For all $u, v \in V$, $u + v = v + u$ ($+$ is commutative);
2. For all $u, v, w \in V$, $(u + v) + w = u + (v + w)$ ($+$ is associative);
3. The set V has a unique zero vector $\bar{0}$.
4. For all $v \in V$ there is a unique element $v' \in V$ such that $v + v' = \bar{0}$.
5. For all $a, b \in F$ and $v \in V$, $a \cdot (b \cdot v) = (ab) \cdot v$;
6. For all $a, b \in F$ and $v \in V$, $(a + b) \cdot v = a \cdot v + b \cdot v$;
7. For all $a \in F$ and $u, v \in V$, $a \cdot (u + v) = a \cdot u + a \cdot v$;
8. For all $v \in V$, $1 \cdot v = v$.

There are several other basic properties that we want. From this set of axioms, however, we can prove these conditions. For example, $0 \cdot v = \bar{0}$ for all $v \in V$. To see this, note that $0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v$ by number 6 above. Letting v' be the additive inverse of v , by adding v' to both sides, we obtain that $\bar{0} = 0 \cdot v + \bar{0} = 0 \cdot v$. Similarly, we can deduce that $a \cdot \bar{0} = \bar{0}$ for all $a \in F$ and $(-1) \cdot v = v'$, that is that $-1 \cdot v$ is the additive inverse of v . Consequently, we actually use $-v$ rather than v' in general.

For our purposes, we will often write 0 for the zero-vector unless it will lead to confusion.

Now suppose we are given two number fields E and F with $F \subseteq E$. Our main tool will be to turn E into a vector space over F . How do we do this? Consider what we really need to make E a vector space over F , an addition operation on E and a “multiplication” operation for elements of F with elements of E . We already have these, namely the field addition of E will be our addition, and the field multiplication of E certainly works as multiplication of an element of F with an element of E .

Theorem 3.5 *If F and E are number fields with $F \subseteq E$, then E is a vector space over F .*

Mathematicians refer to what we are doing as applying a “forgetful” operation. Namely, we are forgetting some of the information given to us. As one mathematician said, “Mathematics is the art of forgetting the right information at the right time.” Why do we want to forget the information? Because vector spaces are a tool that mathematicians know a lot about. In particular, we have many tools with which to study vector spaces, as we shall see.

If E is a vector space over F , a set of vectors $\{v_1, \dots, v_n\} \subset E$ is **linear independent** over F if

$$\sum_{i=1}^n a_i v_i = 0$$

implies $a_i = 0$ for all i . If a set is not linearly independent, we say that it is **linearly dependent**.

We say that a set $\{v_1, \dots, v_n\} \subset E$ **spans** E over F if the set

$$\left\{ \sum_{i=1}^n a_i v_i \mid a_i \in F \right\} = E.$$

The set $B = \{v_1, \dots, v_n\}$ is then a **basis** for E over F if B is a linearly independent spanning set. One can extend these definitions to allow for infinite bases and infinite spanning sets, although we will not do that here. An important theorem of linear algebra then states

Theorem 3.6 *If V is a vector space over F , then there exists a basis for V over F . Moreover, every such basis has the same cardinality.*

We define the **dimension** of E over F to be the cardinality of any basis for E over F , feeling comfortable that this is well defined by the above theorem.

Other ways of defining the dimension that you may have learned in linear algebra (at least in finite cases) is that the dimension is the size of a maximally linearly independent set over F , or a minimal spanning set. The dimension of a vector space is one of the important invariants we shall analyze.

At this point, let us consider an example. We will take the rationals Q for our ground field, and let

$$E = Q[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in Q\}.$$

To see that E is actually a field, we need to check that it is closed under addition, subtraction, multiplication, and non-zero division. Except for division these are all straightforward. For division, however, we just use the process of rationalizing the denominator (this being one of the few times a college professor will have you rationalize a denominator). That is,

$$(a + b\sqrt{2})/(c + d\sqrt{2}) = \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2}.$$

To be sure that this is actually an element of our field E , we need to know that $c^2 - 2d^2 \neq 0$, but if it did, then we would have $\sqrt{2} = c/d$, a rational number, contradicting the irrationality of $\sqrt{2}$. Thus E is a field. What is a basis for E over Q . The obvious basis is the pair $\{1, \sqrt{2}\}$. There are other bases, though. For example $\{1 + \sqrt{2}, 1 - \sqrt{2}\}$ and $\{2 + 7\sqrt{2}, 3 + \sqrt{2}\}$ are also bases for E over Q . In this case, we have that E is 2-dimensional over Q .

When $F \subseteq E$ are number fields, we say that E is a **field extension** of F , and we define the **degree** of E over F to be the dimension of E as a vector space over F . We write these as

$$[E : F] = \dim_F(E).$$

Let's do a couple more examples. This time, let

$$E = Q[\sqrt[3]{2}] := \{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in Q\}.$$

Again, it is fairly easy to check addition, subtraction, and multiplication. This time, it is a lot more complicated to check on how to divide. In particular, what is $1/(3 + \sqrt[3]{2} - 5\sqrt[3]{4})$? In high school, you were never asked to rationalize this denominator, yet you can. (Actually, there was one rather annoying boy in my algebra II class who asked how to rationalize this denominator.) The annoying answer is that you just multiply top and bottom

by $19 + 47\sqrt[3]{2} + 16\sqrt[3]{4}$, and if we do this, we get -381 so that

$$\frac{1}{3 + \sqrt[3]{2} - 5\sqrt[3]{4}} = \frac{-19}{381} + \frac{-47}{381}\sqrt[3]{2} + \frac{-16}{381}\sqrt[3]{4}.$$

That was of course “magic”, and the first time most high school students see it, rationalizing quadratic denominators also seem like magic. (As a brief aside, rationalizing denominators is one of the things that is thankfully leaving the curriculum. It is rarely used and is generally taught as a pure algorithm. The reason why we might have wanted it on the other hand is that it gives students practice on understanding the value of knowing the difference of squares factorization.)

So how did we come up with this magic number above? There are two typical ways to do it. The first you probably learned and promptly forgot in abstract algebra. It involved factor rings of $Q[x]$ by the ideal $(x^3 - 2)$. The second way involves matrix theory. We will do both.

Method 1 (The Euclidean Algorithm for Polynomials): In abstract algebra, you learned that if $p(x)$ is an irreducible polynomial in $Q[x]$ and $(p(x))$ denotes the ideal of all multiples of $p(x)$, then the factor ring $Q[x]/(p(x))$ is isomorphic to $Q[a]$ where a is a root of $p(x)$ (and $Q[a]$ denotes the smallest ring containing the rational numbers and a). The idea is that we define polynomials $f(x)$ and $g(x)$ to be equivalent if $p(x)|(f(x) - g(x))$ (just like we define modular arithmetic). In particular, every polynomial $f(x)$ is equivalent to its remainder $r_f(x)$ upon division by $p(x)$. Thus any polynomial is equivalent to a polynomial of the form $a_0 + a_1x + \dots + a_{n-1}x^{n-1}$, where $\deg(p(x)) = n$. The main theorem you then proved was that if $f(x) \equiv g(x) \pmod{p}$ and $h(x) \equiv k(x) \pmod{p}$, then $f(x) + h(x) \equiv g(x) + k(x)$ and $f(x)h(x) \equiv g(x)k(x)$ modulo $p(x)$ (again, just like in modular arithmetic). Thus $Q[x]/(p(x))$ is a well-defined commutative ring.

The question that arises is when is this a field. Well, since $p(x)$ is irreducible, if $f(x) \in Q[x]$ is not equivalent to the 0 polynomial, then $p(x) \nmid f(x)$, so the greatest common divisor of $p(x)$ and $f(x)$ is 1. Finally, you had a theorem which said that if $d(x)$ was the greatest common divisor of $p(x)$ and $f(x)$, then there existed polynomials $r(x)$ and $s(x)$ such that $p(x)r(x) + f(x)s(x) = d(x)$, and hence in this case, $f(x)s(x) \equiv 1$ modulo $p(x)$, so that $s(x)$ is the “inverse” of $f(x)$. We use the quotation here because, neither $f(x)$ nor $s(x)$ is actually an element of $Q[x]/(p(x))$, but rather

they represent equivalence classes of elements (just like fractions represent equivalence classes of elements of the rationals).

To find $s(x)$, you then enact the Euclidean Algorithm as we are about to do. The difficulty here is that the denominators become truly terrible. In our case above, $p(x) = x^3 - 2$, and the correspondence is that $a + b\sqrt[3]{2} + c\sqrt[3]{4}$ corresponds to the polynomial $a + bx + cx^2$, since x is a root of $p(x)$ in $\mathbb{Q}[x]/(p(x))$. Thus, our case has $f(x) = 3 + x - 5x^2$. Since this case will give very ugly fractions, let us see how everything should work with an easier element. For example, suppose we want to invert $3 + \sqrt[3]{2}$. This corresponds to the function $g_1(x) = x + 3$, and the division algorithm gives:

$$x^3 - 2 = (x^2 - 3x + 9)(x + 3) - 29$$

Thus $29 = -(x^3 - 2) + (x + 3)(x^2 - 3x + 9)$, so that $1 = -\frac{1}{29}(x^3 - 2) + (x + 3)(\frac{1}{29}x^2 - \frac{3}{29}x + \frac{9}{29})$, giving $(3 + \sqrt[3]{2})^{-1}$ as $\frac{9}{29} - \frac{3}{29}\sqrt[3]{2} + \frac{1}{29}\sqrt[3]{4}$. It turns out that despite all appearances, this is a much simpler case than what we started with. Using $g(x) = -5x^2 + x + 3$ now, the first step of the Euclidean algorithm yields

$$x^3 - 2 = \left(\frac{-1}{5}x - \frac{1}{25}\right)(-5x^2 + x + 3) + \left(\frac{16}{25}x + \frac{-47}{25}\right).$$

Of course, the fractions get worse in the next step, at which point we get

$$-5x^2 + x + 3 = \left(\frac{16}{25}x + \frac{-47}{25}\right) * \left(\frac{-125}{16}x - \frac{5475}{256}\right) - \frac{9525}{256}.$$

Of course, now we plug back in to find out what $s(x)$ and $t(x)$ should be so that we get the desired inverse. Clearly this is a job that should be done by a computer, not a human, and resorting to such, we have modulo $x^3 - 2$ that

$$(-5x^2 + x + 3) = \left(\left(\frac{-125}{16}x - \frac{5475}{256}\right)\left(\frac{1}{5}x + \frac{1}{25}\right) - 1\right) * \frac{256}{25 \cdot 381}.$$

When the dust settles, this last is equal to

$$-\frac{16}{381}x^2 - \frac{47}{381}x - \frac{19}{381}.$$

Method 2 (Matrices): Fortunately, there is a way that let's us use our TI-83 calculators, and lets us bypass a lot of the messy calculations. To do this

we use, another way of expressing this field $Q[\sqrt{2}]$ that uses linear algebra. Recall for a vector space V over a field F , a map $f : V \rightarrow V$ is F -linear if

$$f(au + bv) = af(u) + bf(v)$$

for all $a, b \in F$ and $u, v \in V$. It is easy to check (do so!) that if E and F are number fields with $F \subseteq E$ and $\alpha \in E$, then multiplication by α (applying ϕ_α) is an F -linear map of the vector space E over F . Namely, for all $a, b \in F$ and $u, v \in E$, the map $\phi_\alpha(x) = \alpha(x)$ has the property that

$$\phi_\alpha(au + bv) = \alpha(au + bv) = a(\alpha u) + b(\alpha v).$$

Given an F -basis for E , any F -linear map uniquely corresponds to a matrix. It follows that if we choose a basis, ϕ_α can be represented by a unique matrix over F .

Let's do another example. Suppose F is the set of real numbers, and that E is the set of complex numbers. An R -basis for C is $\{1, i\}$. Let the $\alpha = 3 - i$. At this point take a few minutes and try and work out for yourself what the corresponding matrix for α is.

There are two possible answers, depending on whether you like your matrices to multiply on the left or the right. The traditional way is to write the function on the left, so let us do that. Again, let us return to linear algebra. Given a basis $\{v_1, \dots, v_n\}$, and a linear map f , the corresponding matrix for f is $(a_{j,k})$, where $f(v_j) = \sum_{k=1}^n a_{j,k}v_k$. In the case above, $v_1 = 1$, $v_2 = i$, $f(v_1) = f(1) = 3 + (-1)i$, and $f(v_2) = f(i) = 1 + 3i$. Thus $a_{1,1} = 3$, $a_{2,1} = -1$, $a_{1,2} = 1$, and $a_{2,2} = 3$. Thus the corresponding matrix is $\begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}$. You should check that the corresponding matrix for the complex number $a + bi$ is $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Probably in your abstract algebra course, you proved that the complex numbers were isomorphic to the set of real two by two matrices of the above form. This linear algebra reason is why this happens.

So how does this help us with our problem of finding an inverse? In our case, $E = Q[\sqrt[3]{2}]$ had a natural basis (we could choose another, but it would just make matters more confusing) of $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$, and given an element $a + b\sqrt[3]{2} + c\sqrt[3]{4}$, the corresponding matrix is

$$\begin{pmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{pmatrix}.$$

Thus, the element at hand has corresponding matrix

$$\begin{pmatrix} 3 & -10 & 2 \\ 1 & 3 & -10 \\ -5 & 1 & 3 \end{pmatrix}.$$

But the multiplying by the inverse should correspond to the inverse matrix. Thus we can ask our graphing calculator to find the inverse matrix. Now, when it does so, the only problem is that the matrix has decimal entries, but we really wanted fractional entries. The abstract algebra tells us that we can find the decimals as fractions, so it is just a question of how. At this point, we pull a rabbit out of our hat. Well, not really, rather we use one more often unproven fact about linear algebra. The key is that when one finds the inverse by expansion by minors, it turns out that the common denominator is the determinant of the matrix we started with. Thus, if A is the matrix above, we have that $\det(A)A^{-1}$ is a matrix with integer entries. In fact, our calculator tells us it is:

$$\begin{pmatrix} 19 & 32 & 94 \\ 47 & 19 & 32 \\ 16 & 47 & 19 \end{pmatrix}.$$

As the determinant of A is -381 , we get our answer from the first column of A^{-1} , the matrix

$$\frac{1}{381} \begin{pmatrix} -19 & -32 & -94 \\ -47 & -19 & -32 \\ -16 & -47 & -19 \end{pmatrix}.$$

An interesting thing to point out at this time is how linear algebra and linear transformations do so much more for us than expected. In particular, here they have actually told us how to work with fairly complicated expressions of numbers as combinations of roots.

All of this was to help us see why every element of $Q[\sqrt[3]{2}]$ has an inverse. The first method answered the question of why, by telling us that since $x^3 - 2$ was irreducible and $\sqrt[3]{2}$ was a root of $x^3 - 2$, that forced every element to have an inverse. The second portion told us how to find that inverse using the tools we have. We also have a handle on what the dimension of the space $Q[\alpha]$ should be now. It should be the degree of an irreducible polynomial which α is a root of. Thus, $Q[\sqrt[3]{2}]$ is a field extension of degree 3 over Q .

Let us summarize this point in a theorem. First we need to make precise the notation we have been using. If $\alpha \in C$ (where C is the set of complex

numbers) and K is subfield of C (note, just think that $K = Q$ for now), then by $K[\alpha]$ we mean the smallest subring of C containing K and α . Here we have

$$K[\alpha] = \left\{ \sum_{I=0}^k a_i \alpha^I \mid a_i \in K \text{ and } k \in \mathbb{Z} \right\}$$

By $K(\alpha)$, we mean the smallest subfield of C containing K and α .

Theorem 3.7 *If $f(x)$ is an irreducible polynomial with coefficients in a number field K , and α is a root of $f(x)$ in C , then $K(\alpha) = K[\alpha]$, and the degree of the field extension $K[\alpha]$ over K is the degree of $f(x)$.*

Proof: From abstract algebra, we have $K[x]/(f(x)) \cong K[\alpha]$. But every polynomial of $K[x]$ is equivalent modulo $f(x)$ to a polynomial of degree less than n by the division algorithm. Thus $\{1 + (f(x)), x + (f(x)), \dots, x^{n-1} + (f(x))\}$ spans $K[x]/(f(x))$ over K . Any non-zero polynomial $g(x) = \sum_{i=0}^{n-1} a_i x^i$ is not divisible by $f(x)$, so that $g(x) + (f(x)) \neq 0 + (f(x))$. Consequently, $\{1 + (f(x)), \dots, x^{n-1} + (f(x))\}$ is linearly independent, and the dimension of $K[x]/(f(x))$ over K is n .

Q.E.D.

We are now ready to state our “big” theorem for this section. The dimensions multiply theorem.

Theorem 3.8 *If F and E are number fields with $Q \subseteq F \subseteq E$, then*

$$\dim_Q(E) = \dim_Q(F) \cdot \dim_F(E).$$

Proof: Let $\{\alpha_1, \dots, \alpha_n\}$ be a basis for F over Q , and let $\{\beta_1, \dots, \beta_m\}$ be a basis for E over F . We will show that

$$B = \{\alpha_i \beta_j \mid i \in \{1, \dots, n\}, j \in \{1, \dots, m\}\}$$

is a basis for E over Q . We begin by showing that B is a linearly independent set. Suppose

$$\sum_{i,j} a_{ij} \alpha_i \beta_j = 0,$$

where the a_{ij} 's are all rational numbers. Rearranging the terms in the summation, we have that

$$0 = \sum_{i=1}^n \sum_{j=1}^m a_{ij} \alpha_i \beta_j \quad (3.1)$$

$$= \sum_{j=1}^m \left(\sum_{i=1}^n a_{ij} \alpha_i \right) \beta_j. \quad (3.2)$$

Now, the set of β_j is linearly independent over F , and since $\sum_{i=1}^n a_{ij} \alpha_i$ is an element of F as the α_i 's are all in F , we have by the definition of linearly independence that

$$\sum_{i=1}^n a_{ij} \alpha_i = 0$$

for all j . Since the set of α_i 's is linearly independent over the rationals, and the a_{ij} 's are all rational, we then have that $a_{ij} = 0$ for all i and j . Hence the set B is linearly independent over Q .

We now must check that B is a spanning set for E over Q . Let x be an arbitrary element of E . Since the set of β_j 's span E over F , there exist elements $b_j \in F$ such that $x = \sum_{j=1}^m b_j \beta_j$. Since the b_j 's are in F and the α_i 's span F over Q , there exist $a_{ij} \in Q$ such that $b_j = \sum_{i=1}^n a_{ij} \alpha_i$ for all j . Unraveling the above equations, we have that

$$x = \sum_{j=1}^m b_j \beta_j \quad (3.3)$$

$$= \sum_{j=1}^m \left(\sum_{i=1}^n a_{ij} \alpha_i \right) \beta_j \quad (3.4)$$

$$= \sum_{i,j} a_{ij} \alpha_i \beta_j. \quad (3.5)$$

Since x was an arbitrary element of E , we have that B spans E over Q . As we have shown that B is a linearly independent spanning set for E over Q , we have shown that B is a basis for E over Q . Since $|B| = nm$, we have shown that $\dim_Q(E) = \dim_Q(F) \cdot \dim_F(E)$.

Q.E.D.

In fact, this theorem is more general than stated. We didn't really need for the smallest field to be the rational numbers. It could have been any subfield of F .

Teaching Aside

At this point, you are probably wondering how any of this really ties into the high school curriculum. In a basic way, the division algorithm for polynomials and elementary matrix theory are in the curriculum so that these points show a natural extension. On the other hand, we claim that there is more to it than that.

In particular, one of the great question that many students ask when you do long division of polynomials, is why should we do this? If you take a utility standpoint, the answer lies in the division algorithm and the Euclidean algorithm (our first technique for inverting polynomials). Most of the error-correcting coding systems, use polynomial arithmetic extensively. In truth, this arithmetic is usually done over a finite field rather than Q , but these systems require the division algorithm and the Euclidean algorithm. Similarly, most secret coding systems use the Euclidean algorithm for integers.

What about problem solving? Let's think about how we solved the problem of finding an inverse of $3 + \sqrt[3]{2} - 5\sqrt[3]{4}$. In both cases, we turned the problem into a completely different problem that used ideas that we already knew. This is an important problem solving strategy that needs to be taught. Moreover, matrices and linear algebra are one of the strongest tools to use in this way. That is, the applications of matrices are wide ranging and help us solve many different problems.

The last issue that we would like to raise here is that the idea of multiple representations is an important strand running through the NCTM standards. What we have seen here is that there are three very different ways to think about something as simple as the $\sqrt[3]{2}$, each with its own strengths and weaknesses. Certainly, in a high school classroom, one can remember these different representations and answer the sort of questions that might require using one of them.

End of Aside

3.4 Impossibility Theorems

The impossibility theorems for the constructibility of $\sqrt[3]{2}$ and trisecting the general angle are very similar. The key to both is theorem 3.8 from the previous section. We shall show that any time you construct a new point on the plane using straightedge and compass, the numerical values for the

coordinates of this new point only involve the numerical values for the coordinates of points already known and the square roots of such values. In the language above, the coordinates lie in a field of dimension 2 over the smallest field containing all previously constructed coordinates. Consequently, by Theorem 3.8, every field extension E obtained via straightedge and compass construction has dimension 2^n over Q . But suppose $\sqrt[3]{2}$ was constructible, then $Q[\sqrt[3]{2}]$ would be a subfield of one of these fields E . But then Theorem 3.8 would imply that $3|2^n$, a contradiction. Thus we cannot construct $\sqrt[3]{2}$. For the trisection, we do something similar. Namely, we show that if you can construct a 20 degree angle, then you would be able to construct a root of an irreducible cubic polynomial with rational coefficients. Again, this would imply that one of our fields E of dimension 2^n would contain a subfield of dimension 3 over Q , a contradiction.

Now that we have outlined our attack, let's fill in the gaps a little better. The first thing we need to do is to answer the following three questions:

1. Suppose you have four points $A = (x_1, y_1)$, $B = (x_2, y_2)$, $C = (x_3, y_3)$, and $D = (x_4, y_4)$, with all coordinates lying in a field E , and \overline{AB} and \overline{CD} intersect. What are the coordinates of the intersection point?
2. Suppose you have the same four points, and the line \overline{AB} intersects the circle with center C through point D . What are the coordinates of the intersection point?
3. Now suppose the circle with center at A through B intersects the circle with center at C through D , what are the coordinates of the intersection point?

In all cases, you should have found out that the coordinates involve at most the square root of some number from the field E . (This is in-class group work and homework.)

Let us summarize what you just showed in a lemma.

Lemma 3.9 *Let E be a subfield of the real numbers, and suppose*

$$x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4 \in E.$$

If (x, y) is a point

1. *at the intersection of the line joining (x_1, y_1) and (x_2, y_2) and the line joining (x_3, y_3) and (x_4, y_4) , then $x, y \in E$.*

2. at the intersection of the line joining (x_1, y_1) and (x_2, y_2) and the circle with center (x_3, y_3) and through (x_4, y_4) , then there exists an element $d \in E$ such that $x, y \in E[\sqrt{d}]$.
3. at the intersection of the circle centered at (x_1, y_1) through (x_2, y_2) and the circle centered at (x_3, y_3) through (x_4, y_4) , then there exists $d \in E$ such that $x, y \in E[\sqrt{d}]$.

At this point let us finish the proof of the impossibility theorems. To begin with, we now need to be precise about our terminology. We follow Stewart [16]. Assume that P_0 is a set of points in the Euclidean plane R^2 , and consider operations of two kinds:

1. through any two points draw a straight line, and
2. Draw a circle centered at one point and through any other point.

A point $(x, y) \in R^2$ is said to be *constructible* from P_0 , if there exist a sequence of points

$$(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t) = (x, y)$$

such that the point (x_i, y_i) can be obtained as the intersection point of two distinct lines or a line and a circle or two distinct circles when drawn using one of our two operations from the points $P_0 \cup \{(x_1, y_1), \dots, (x_{i-1}, y_{i-1})\}$. We say that (x, y) is *constructible over the rationals* if (x, y) is constructible over the set $P = \{(a, b) \mid a, b \in Q\}$.

We will prove our result in a series of lemmas.

Lemma 3.10 *Suppose (x, y) is constructible over the rationals via the sequence $(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t) = (x, y)$. Define $K_i = Q[x_1, y_1, \dots, x_i, y_i]$ to be the smallest field containing Q and the numbers $x_1, y_1, x_2, \dots, y_i$. Then $K_i = K_{i-1}[\sqrt{a}]$ for some $a \in K_{i-1}$.*

Proof: This follows from Lemma 3.9.
Q.E.D.

Lemma 3.11 *Using the notation of the preceding Lemma, we have $\dim_{K_{i-1}} K_i$ is 1 or 2.*

Proof: By the preceding lemma, $K_i = K_{i-1}[\sqrt[a]{a}]$ for some $a \in K_{i-1}$. Thus, $x^2 - a$ is either an irreducible polynomial over K_{i-1} or $\sqrt{a} \in K_{i-1}$. By Theorem 3.7, we have $\dim_{K_{i-1}} K_i$ is either 1 or 2. Q.E.D.

We now define a real number α to be *constructible* if $(\alpha, 0)$ is constructible over Q .

Theorem 3.12 *Suppose x is a constructible number. Then $[Q(x) : Q]$ is a power of 2.*

Proof: As x is constructible, there exists a sequence of points

$$(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t) = (x, 0)$$

as in the definition of a constructible point $(x, 0)$. Let K_i be defined as before. Then $x \in K_t$. We now prove by induction on i that K_i has dimension a power of 2 over Q . Since $K_0 = Q$, it has dimension $1 = 2^0$ over Q establishing the basis step. Suppose $i > 0$ and that K_{i-1} has dimension 2^l over Q for some l . By Lemma 3.11, we have $\dim_{K_{i-1}} K_i$ is 1 or 2. Consequently, by the degrees multiply theorem, it follows that $\dim_Q(K_i)$ is equal to 2^l times one or two, and hence is again a power of two. By induction, $\dim_Q(K_t) = 2^l$ for some l .

As $x \in K_t$, it follows that $Q(x)$ is a subfield of K_t . The degrees multiply theorem then implies that $\dim_Q(K_t) = \dim_Q(Q(x)) \cdot \dim_{Q(x)}(K_t)$. Thus $\dim_Q(Q(x))$ divides $\dim_Q(K_t) = 2^l$. Thus $\dim_Q(Q(x)) = [Q(x) : Q]$ must also be a power of 2.

Q.E.D.

Theorem 3.13 *The $\sqrt[3]{2}$ is not constructible.*

Proof: We have already seen that $[Q(\sqrt[3]{2}) : Q] = 3$. Thus Theorem 3.12 implies $\sqrt[3]{2}$ is not constructible.

Q.E.D.

Question: Can you find an irreducible polynomial satisfied by $\cos(20^\circ)$? Think about how one can use the angle sum formulas and the knowledge that $\cos(60^\circ) = .5$ to find an equation that $\cos(20^\circ)$ is a root of. (In class homework assignment).

Based on your answer to the last problem, you should have discovered that $\cos(20)$ is the root of an irreducible cubic polynomial with integer coefficients. This implies by our previous arguments that $\cos(20)$ cannot be constructed.

Theorem 3.14 *There does not exist an algorithm which can trisect a 60° angle with straightedge and compass.*

Proof: Suppose there did exist such an algorithm. As an equilateral triangle can be constructed, we can construct a 60° angle. Consequently, we could construct a 20° angle. From this it follows that we can construct a line making a 20° angle with the x -axis of length 1 from the origin. Thus, as in the picture below, by dropping a perpendicular, we can construct the

point $(\cos(20), 0)$, implying that $\cos(20)$ is constructible. Having achieved a contradiction, we see that we cannot trisect a 60° angle. Q.E.D.

3.5 Regular n -gons

One of the great feats of Gauss was to discover with proof, which of the regular n -gons can be constructed with straightedge and compass. Unfortunately, the proof of Gauss's result goes beyond the scope of this book, but let us, but let us at least state the result.

Theorem 3.15 *A regular n -gon can be constructed with straightedge and compass if and only if $n = 2^{k_1} p_1 p_2 \dots p_t$, where k_1 is a non-negative integer, and the p_i s are distinct Fermat primes.*

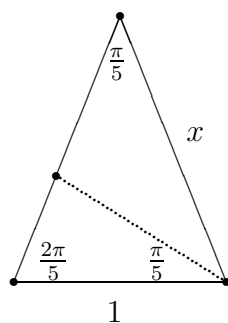
A Fermat prime is a prime number of the form $2^{2^s} + 1$, so that 3, 5, 17, 257, and 65537 are the first four Fermat primes. Thus, a regular 17-gon can be constructed, but a regular 9-gon cannot. Briefly, the proof of this theorem seems to require Galois theory or mathematics at roughly the same level, and the interested reader is encouraged to look at [16].

It turns out that we have already handled the question of the regular 9-gon since constructing such a figure would require constructing a 40° exterior angle which could then be bisected to produce a 20° angle, a contradiction.

However, while we cannot prove Gauss's theorem (and the 17-gon is extremely complicated to construct), it does make sense to stop here and work on constructing a regular pentagon.

So, how can you do such a thing? We now have a great deal of information about constructing numbers. We know how to construct sums and products, and earlier we found out how to construct \sqrt{a} given a , so we shall use these ideas to construct our pentagon. To do so, work through the following set of questions.

- C1.** What is the exterior angle of a regular 5-gon. That is, what angle do you need to construct to construct the 5-gon.
- C2.** Consider the following figure



What is the measure of x ? Show this. How can you use this to construct the needed angle.

- C3.** Construct a segment of length x .
- C4.** Construct an isosceles triangle ACD with two sides (AC and AD) of length x and one side CD of length 1. Construct points B and E at distance 1 from A and respectively of distance 1 from C and D .
- C5.** Argue that $ABCDE$ is a regular pentagon.

3.6 Problems

Warm up problems

1. Give an example of two pentagons which have corresponding angles congruent but are not similar.
2. Suppose V is a vector space over the field F . Prove for all $a \in F$ that $a \cdot \bar{0} = \bar{0}$.
3. Construct a regular pentagon with straightedge and compass. Justify the steps along the way.
4. Given segments of length a and b and unit length 1, show how to construct ab and $\frac{a}{b}$. Prove that these constructions give the appropriate numbers.
5. In the text we have given one way to construct the square root of a given integer x by using a right triangle. Give a second proof by using mathematical induction on $x \geq 2$ and the Pythagorean Theorem.
6. Without actually doing the construction, show that

$$\sqrt{2 + \sqrt[4]{3 + \sqrt{2}}}$$

is constructible.

7. Show that $\sqrt[3]{25}$ cannot be written as $a + b\sqrt[3]{5}$ for any rational numbers a and b .
8. The complex numbers C can be viewed as a two-dimensional vector space over the real numbers. The standard basis for C is the set $\{1, i\}$, and the representation of the complex number $a + bi$ as a 2×2 matrix using this basis is

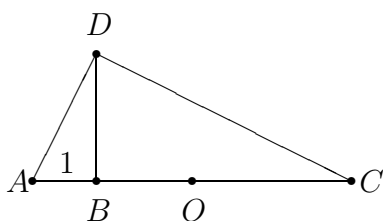
$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Find the determinant of this matrix. The square root of this determinant is the *norm* of the complex number $a + bi$. What is the matrix associated to the element $a + bi$ using the basis $\{1 + i, 1 - i\}$? What relationship does this determinant have with the norm of the complex number? Explain.

9. Is $\sqrt[5]{45}$ constructible.
10. Given a triangle provide the constructions for inscribing a circle within the triangle and circumscribing a circle about the triangle.

Advanced Problems

11. In the next problem, we show how to construct \sqrt{a} given a unit length and a length a . To do this we will use the theorem from Euclidean geometry which states: if an angle A is inscribed in a circle, then the arc cut off by A has measurement $2|A|$ where $|A|$ denotes the radian measure of A . Prove this.
12. Given the following picture, prove that $|DB| = \sqrt{|CB|}$, given that the circle centered at O through A , also runs through C and D .



13. Without actually doing the construction show that

$$\left(\sqrt[4]{3 + \sqrt{7} + \sqrt{5}}\right) \left(\sqrt[8]{\sqrt{3} + \sqrt{5} + 2}\right)$$

is constructible.

14. Suppose a line defined by two points with coordinates in a field F is intersected with a circle having center with coordinates in F and radius in F . Prove that the points of intersection have coordinates in a field F' which is two dimensional over F .
15. Prove without working too hard that $Q[\sqrt{2}, \sqrt[3]{2}]$ is a field of dimension 6 over the rational numbers.

16. Find the multiplicative inverse of the number $x = 1 + 2\sqrt[3]{3} + 2\sqrt[3]{9}$, and prove that your answer is correct.
Hint: To do this, you simply need to come up with a candidate for the inverse and multiply the two numbers together and see that you get 1. To find a candidate you should follow the steps given:
- (a) Using the basis $\{1, \sqrt[3]{3}, \sqrt[3]{9}\}$ for the field $Q[\sqrt[3]{3}]$ over the field Q , write the 3×3 matrix M corresponding x viewed as a linear transformation from Q^3 to Q^3 .
 - (b) Use Gaussian elimination to find the inverse of the matrix M .
 - (c) Find the element y of $Q[\sqrt[3]{3}]$ corresponding to M . Now multiply x and y to check that $xy = 1$.
17. If we take the basis of $\{1, \sqrt{3}, \sqrt{5}, \sqrt{15}\}$ for the field $Q[\sqrt{3}, \sqrt{5}]$ over Q , then what is the matrix associated to the element $2 + \sqrt{3} - \sqrt{15}$ of this ring? What is the inverse of this element?
18. Let n be an integer. Find necessary and sufficient conditions that an angle of n degrees can be constructed with straightedge and compass. (Hint, show that a 3° angle can be constructed and that a 2° angle cannot be constructed with straightedge and compass.)

Chapter 4

Solving Equations by Radicals

In the last chapter we proved that $\sqrt[3]{2}$ is not a constructible number. Obviously, we might do somewhat better if we used different tools of construction. In fact, the ancient Greeks knew how to trisect the general angle if they were allowed to use extra tools. Similarly, they could construct cube roots using other tools. Today, one of the methods used to do constructive type geometry in the schools is by allowing the construction tools that you have when paper folding. It has been shown that using paper folding, it is possible to construct $\sqrt[3]{2}$. Nevertheless, for each of these expanded definitions of construction tools, a proof similar to the one given in the last chapter will show that there exists some numbers that cannot be constructed. In fact, in all of these cases, the proof shows that there is some n th root of 2 that cannot be constructed.

On the other hand, we have an algebraic definition for the number $\sqrt[n]{2}$. Thus, we can ask, can we get all real numbers via combinations of *radicals*. By this, we mean is it the case that any real number can be written using just the operations of addition, subtraction, multiplication, division, and taking an n th root, together with the rational numbers? It is certainly true that we can get a lot of numbers this way that we couldn't get via constructions. For example,

$$\sqrt[7]{\frac{2}{3} - \sqrt[9]{5 + \sqrt[3]{2}}}$$

is one such number. Can this be everything? As a simpler question, we might ask simply, can we solve every algebraic equation that has real roots this way? (We could even go so far as to ask the question for complex

roots since after all, $i = \sqrt{-1}$). While on the face of it, this question might look straightforward, many great mathematicians had difficulty with this question. For example, Euler wrote at least one article in which he suggested that this should be the case, but there was a mistake found in his work.

Since we know that every linear equation with rational coefficients can be solved over the rational numbers, the first interesting case for us will be the quadratic equation. You should start by working through the following problem:

Derive the quadratic formula.

Chances are that your derivation looks something like the following:

Derivation of the Quadratic Formula: Suppose $ax^2 + bx + c = 0$ where a , b , and c are real numbers with $a \neq 0$. Subtracting c from both sides, this is equivalent to

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

Dividing both sides of the equation by a , our original equation is then equivalent to

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

At this point, we add $(\frac{b}{2a})^2$ to both sides so that the left hand side of the equation is a perfect square. Thus x solves our original equation if and only if x solves

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

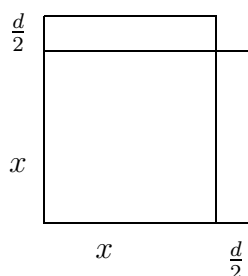
The right hand side of this equation is then $(x + \frac{b}{2a})^2$, so that

$$x + \frac{b}{2a} = \pm \sqrt{\frac{-c}{a} + (\frac{b}{2a})^2}.$$

Thus x is a root of our original equation if and only if

$$\begin{aligned} x &= \frac{-b}{2a} \pm \sqrt{\frac{-c}{a} + (\frac{b}{2a})^2} \\ &= \frac{-b}{2a} \pm \sqrt{\frac{-4ac}{(2a)^2} + \frac{b^2}{(2a)^2}} \\ &= \frac{-b}{2a} \pm \frac{1}{2a} \sqrt{-4ac + b^2} \\ &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \end{aligned}$$

The above derivation is relatively easy to do **if you are good at algebra**. Of course, if you aren't, it is extremely hard to recreate and looks pretty much like random symbols. Also, while the notion of “completing the square” is algebraically clear when you add in the $\frac{b^2}{4a^2}$ term, the idea doesn't stick for most students. One can also derive the quadratic formula geometrically, but for this, let us consider the special case where the equation we are trying to solve is $x^2 + dx = e$, where d and e are non-negative. Then we can geometrically represent $x^2 + dx$ as the area of the rectangle below:



Thus in the picture, to complete the square, you need to add a little square of side length $\frac{d}{2}$. Thus, if the equation were read as $x^2 + dx = e$ with d and e positive, the thinking would be that you would need to add a little piece of area $(\frac{d}{2})^2$ to get a square of area $e + (\frac{d}{2})^2$. Thus, the side length of the square must be $\sqrt{e + (\frac{d}{2})^2}$. Moving the d over to the other side, we then get that $x = -\frac{d}{2} + \sqrt{e + (\frac{d}{2})^2}$. Of course, this isn't the usual statement of the quadratic formula. However, as a first introduction to the formula, this can be very helpful, as it grounds the formula in a geometric notion that many students feel more comfortable with and can recreate at a later point. Moreover, while we have assumed that d and e are positive, the algebraic justification turns out to be equally good, even if they are negative, so that this derivation can give students a way to recreate the formula.

To turn this into the usual version of the quadratic formula is just a matter of substitution. That is, suppose we are to solve the equation $ax^2 + bx + c = 0$ where $a \neq 0$. Dividing through by a and subtracting $\frac{c}{a}$ from each side we obtain the equivalent equation

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

Thus, the equation is the same as earlier with $d = \frac{b}{a}$ and $e = \frac{-c}{a}$. Conse-

quently, our earlier solution tells us that

$$\begin{aligned}
 x &= -\frac{d}{2} \pm \sqrt{e + \left(\frac{d}{2}\right)^2} \\
 &= -\frac{b}{2a} \pm \sqrt{\frac{-c}{a} + \left(\frac{b}{2a}\right)^2} \\
 &= -\frac{b}{2a} \pm \sqrt{\frac{-c}{a} + \left(\frac{b^2}{4a^2}\right)} \\
 &= -\frac{b}{2a} \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} \\
 &= -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}
 \end{aligned}$$

as desired.

Of course, when solving a quadratic equation with this approach, it might make more sense (at least initially) to solve it step by step rather than apply the formula. For example, given the equation $2x^2 + 5x - 8 = 0$, the first step is to divide the equation through by 2, yielding $x^2 + \frac{5}{2}x - 4 = 0$. At this point we can apply our reduced quadratic formula with $d = \frac{5}{2}$ and $e = 4$, yielding

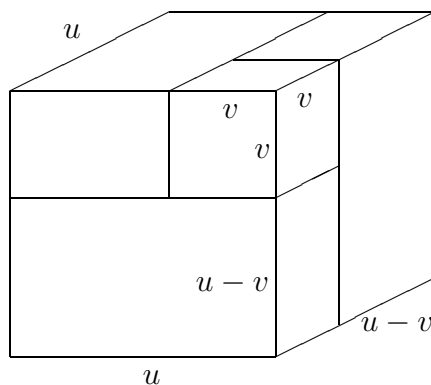
$$\begin{aligned}
 x &= -\frac{5}{4} \pm \sqrt{4 + \left(\frac{5}{4}\right)^2} \\
 &= -\frac{5}{4} \pm \sqrt{4 + \frac{25}{16}} \\
 &= -\frac{5}{4} \pm \sqrt{89/16} \\
 &= \frac{-5 \pm \sqrt{89}}{4}.
 \end{aligned}$$

In truth, one can skip the last three lines here as the first statement is correct, it just isn't simplified in the standard way. Of course, depending on what you want the answer for, it doesn't need to be simplified.

4.1 Solving Simple Cubic Equations

Now, let us think about these geometric methods and what they might mean in three dimensions. Consider a cube of side length u , with a little cube cut

out of the corner having side length v . The volume of this figure is then $u^3 - v^3$. Breaking the shape into 4 pieces as in the figure, however, we obtain a different formula for the volume, namely $(u - v)^3 + 3uv(u - v)$. At this



point, we have the formula

$$(u - v)^3 + 3uv(u - v) = u^3 - v^3.$$

Suppose we let $x = (u - v)$, then this formula becomes $x^3 + (3uv)x = u^3 - v^3$. Consequently, we might be able to use this to solve cubic polynomials of certain types.

Suppose we are to solve the equation $x^3 + px = q$ where p and q are both positive. If we want to use the above to attempt to solve this problem, we would let $x = u - v$, and we would want to find u and v such that $3uv = p$ and $u^3 - v^3 = q$. Solving the first of these equations for v , we obtain $v = \frac{p}{3u}$. Substituting this in for v in the second equation, we obtain

$$u^3 - \left(\frac{p}{3u}\right)^3 = q.$$

Multiplying this equation through by u^3 , we then obtain

$$u^6 - \left(\frac{p}{3}\right)^3 = qu^3.$$

This equation becomes a quadratic in u^3 , yielding

$$(u^3)^2 - q(u^3) - \left(\frac{p}{3}\right)^3 = 0.$$

We can now use the quadratic formula to obtain a value for u^3 . Namely,

$$u^3 = \frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

Hence

$$u = \sqrt[3]{\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

We can now plug the value for u^3 into the equation $u^3 - v^3 = q$ to obtain that

$$\begin{aligned} v^3 &= \frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} - q \\ &= -\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}. \end{aligned}$$

Thus, taking the cube root, we obtain that

$$v = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

As the quantity under the square root sign is necessarily positive, once we choose which of the square roots we use for u , we get a unique real solution. Suppose for the time being that for u we take the positive square root. As $x = u - v$, we end up with the solution:

$$x = \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} - \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Curiously, as you will show in homework ??, choosing the other sign for u yields the same answer for x . This technique gives us at least one solution to the equation in the case where p and q are both non-negative.

Example 1: Consider the equation $x^3 + 2x = 8$. In this case, $p = 2$ and $q = 8$. Using our formula, we have

$$x = \sqrt[3]{4 + \sqrt{4^2 + \left(\frac{2}{3}\right)^3}} - \sqrt[3]{-4 + \sqrt{4^2 + \left(\frac{2}{3}\right)^3}}$$

$$\begin{aligned}
&= \sqrt[3]{4 + \sqrt{16 + \frac{8}{27}}} - \sqrt[3]{-4 + \sqrt{16 + \frac{8}{27}}} \\
&= \sqrt[3]{4 + 2\sqrt{4 + \frac{2}{27}}} - \sqrt[3]{-4 + 2\sqrt{4 + \frac{2}{27}}}.
\end{aligned}$$

Example 2: Consider the equation $x^3 + 3x = 14$. In this case, $p = 3$ and $q = 14$. For the given solution above, we have $\frac{p}{3} = 1$ and $\frac{q}{2} = 7$. Thus a real solution is given by

$$x = \sqrt[3]{7 + \sqrt{7^2 + 1^3}} - \sqrt[3]{-7 + \sqrt{7^2 + 1^3}}.$$

Simplifying this expression, we see that

$$x = \sqrt[3]{7 + 5\sqrt{2}} - \sqrt[3]{-7 + 5\sqrt{2}}.$$

Before going on and trying to pull out a factor from our equation above so that we can use the quadratic formula, let us take a minute and see if we can simplify this term at all. If we plug this into our calculator, we find that it is so close to 2 that our calculator thinks it is 2. So, we have to ask, is it 2?

We have actually discussed a question like this before. If it is 2, then we would know that $\{1, \sqrt[3]{7 + \sqrt{7^2 + 1^3}}, \sqrt[3]{-7 + \sqrt{7^2 + 1^3}}\}$ is a linearly dependent set over \mathbb{Q} . (Of course, if it isn't 2, that does not establish that the set is independent.) Now that we know we have seen the question before, does that help us solve it? Well, not really. The difficulty is that we need to find out whether there is an easier way of writing the cube root of $7 + 5\sqrt{2}$. We might guess that the cube root should have the form $a + b\sqrt{2}$ for two rational numbers a and b . If so, then we would know that $(a + b\sqrt{2})^3 = 7 + 5\sqrt{2}$. On the bright side, we can actually cube the left hand side. Doing this, we obtain

$$\begin{aligned}
7 + 5\sqrt{2} &= a^3 + 3a^2b\sqrt{2} + 3a(b\sqrt{2})^2 + (b\sqrt{2})^3 \\
&= a^3 + 3a^2b\sqrt{2} + 3ab^2 \cdot 2 + 2b^3\sqrt{2} \\
&= a^3 + 6ab^2 + (3ab^2 + 2b^3)\sqrt{2}.
\end{aligned}$$

Since we know that $\{1, \sqrt{2}\}$ is linearly independent over the rational numbers, the only way we can have $7 + 5\sqrt{2} = a^3 + 6ab^2 + (3ab^2 + 2b^3)\sqrt{2}$ is if

$$\begin{aligned}
a^3 + 6ab^2 &= 7 & \text{and} \\
3a^2b + 2b^3 &= 5.
\end{aligned}$$

If we are really lucky, we can solve this by inspection, and in this case we are quite fortunate in that $a = 1$ and $b = 1$ is indeed a solution. Thus

$$\sqrt[3]{7 + 5\sqrt{2}} = 1 + \sqrt{2}.$$

A similar check shows that

$$\sqrt[3]{-7 + 5\sqrt{2}} = -1 + \sqrt{2},$$

so that

$$\begin{aligned} x &= \sqrt[3]{7 + 5\sqrt{2}} - \sqrt[3]{-7 + 5\sqrt{2}} \\ &= (1 + \sqrt{2}) - (-1 + \sqrt{2}) \\ &= 2, \end{aligned}$$

as we suspected.

This example has shown us that writing numbers with radicals is fraught with dangers. Numbers that don't even look like rational numbers might turn out to be rational numbers. We could have seen this just using square roots since $\sqrt{3 + 2\sqrt{2}} = 1 + \sqrt{2}$, but the example above actually arises quite naturally from our solution to the cubic equation. In fact, if you look at the original equation, you can immediately see that $x = 2$ is a solution.

While we initially assumed p and q are both positive, we only needed to do this because we were talking about lengths and volumes. However, the algebraic statement $u^3 - v^3 = 3uv(u - v) + (u - v)^3$ is true independent of the volume argument. Consequently, we need not assume that both u and v are positive for the argument to make sense. Consequently, we can try and do everything the same way, even when p and q are not necessarily positive. So what might go wrong in this case? A difficulty might arise because the quantity under the square root sign might be negative, leading to an imaginary square root, which we don't know how to take the cube root of. For example, consider the equation

$$x^3 - 15x = 4$$

For our solution, we have $p = -15$ and $q = 4$. Thus $\frac{p}{3} = -5$ and $\frac{q}{2} = 2$. Plugging these values into the equation, we obtain:

$$\begin{aligned} x &= \sqrt[3]{2 + \sqrt{2^2 - 5^3}} - \sqrt[3]{-2 + \sqrt{2^2 - 5^3}} \\ &= \sqrt[3]{2 + \sqrt{-121}} - \sqrt[3]{-2 + \sqrt{-121}}. \end{aligned}$$

In this case, it is not at all clear what to do. However, if we look at the original equation, $x = 4$ is a solution! This was, in fact, one of the equations that had the Italian mathematicians stymied.

The Italian mathematician Bombelli had a “wild” idea. He thought to treat $\sqrt{-1}$ as just another algebraic symbol. Working as we did above on taking the cube root of $7 + 5\sqrt{2}$, he attempted to find numbers a and b so that

$$(a + b\sqrt{-1})^3 = 2 + \sqrt{-121}.$$

Setting $\sqrt{-121} = 11\sqrt{-1}$, Bombelli cubed the left hand side to get

$$a^3 - 3ab^2 + (3a^2b - b^3)\sqrt{-1} = 2 + \sqrt{-121}.$$

Thus, he needed to find values for a and b such that

$$\begin{aligned} a^3 - 3ab^2 &= 2 \quad \text{and} \\ 3a^2b - b^3 &= 11. \end{aligned}$$

One solution is given by $a = 2$ and $b = 1$. Thus $\sqrt[3]{2 + \sqrt{-121}} = 2 + \sqrt{-1}$. The cube root of $-2 + \sqrt{-121}$ turns out to be $-2 + \sqrt{-1}$, so that the Cardano method of solving the cubic yields

$$x = 2 + \sqrt{-1} - (-2 + \sqrt{-1}) = 4$$

as we expected.

Teaching and Historical Aside:

Many textbooks assert that the complex numbers arose in mathematics to solve the equation $x^2 + 1 = 0$, which is simply not true. In fact, they originally arose as in the above example. The complex numbers were certainly not accepted as anything more than a useful tool for centuries. A century ago, the great mathematician Felix Klein gave the following description of the history of the complex numbers [5]

...imaginary numbers made their own way into arithmetic calculation without the approval, and even against the desires of individual mathematicians, and obtained wider circulation only gradually and to the extent to which they showed themselves useful. Meanwhile the mathematicians were not altogether happy about it. Imaginary numbers long retained a somewhat mystic

coloring, just as they have today for every pupil who hears for the first time about that remarkable $i = \sqrt{-1}$. As evidence, I mention a very significant utterance by Leibniz in the year 1702, “Imaginary numbers are a fine and wonderful refuge of the divine spirit, almost an amphibian between being and non-being.” In the eighteenth century, the notion was indeed by no means cleared up, although Euler, above all, *recognized their fundamental significance for the theory of functions*. In 1748 Euler set up that remarkable relation:

$$e^{ix} = \cos x + i \sin x$$

by means of which one recognizes the fundamental relationship among the kinds of functions which appear in elementary analysis. The *nineteenth century finally brought the clear understanding of the nature of the complex numbers*.

What Klein doesn't mention is the difficulty that many great mathematicians had dealing with the complex numbers. In one paper, Leibniz argued that \sqrt{i} was not a complex number, apparently not recognizing that $(\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}})^2 = i$. This difficulty was a reflection of the difficulty mathematicians had earlier with the concept of negative numbers.

Today, the complex numbers show up in physics problems, and do have a geometric meaning as we shall explore below, in an attempt to find all of the solutions to our cubic. Thus, when teaching about the complex numbers, it is important to remember that we work with them because they are useful, and that they are a natural extension of the idea of number.

End of Aside

At this point, we have a way to find real roots for some equations. Before turning to the question of finding all roots, which will require investigating the complex plane, let us first show how to use the special case solution to solve the more general cubic.

4.2 The General Cubic Equation

Suppose we are given the equation

$$x^3 + bx^2 + cx + d = 0.$$

We would like to use our previous work to solve this problem, but the difficulty is that this time we have a quadratic term (if $b \neq 0$). Consequently, we would like to find a way to eliminate this term. At this point, let us return to the quadratic equation and examine it again to see if we can find another way to approach the general cubic.

Consider the function $f(x) = x^2 + bx + c$. Solving a quadratic equation is equivalent to finding the roots of the function $f(x)$. Rather than transforming the problem its equation form, however, let us examine the function form more. Certainly, the roots of the function $g(x) = x^2 + c$ can be easily found to be $x = \pm\sqrt{-c}$. Thus, it might be helpful to find a way to transform the function $f(x)$ into a function of the form $g(x)$. How might we do this? To come up with an answer, requires that we look at the properties of the graph of $g(x)$. The key element here is that the graph of $g(x)$ is symmetric about the y -axis. Thus, we would like to transform $f(x)$ into a function that is symmetric about the y -axis. The graph of $f(x)$ is a parabola, and any parabola is symmetric about a line through its vertex. Consequently, if we translate the graph of $f(x)$ sufficiently, we should be able to arrive at a case where it too is symmetric about the y -axis. But what should be translate the function by. There are several options here. One is to calculate $f(x - d)$ and discover what value of d eliminates the linear term, another would be to guess and check, and a third would be to use calculus to find the x -value of the vertex (where $f'(x) = 0$). Working through any one of these ideas, we see that the appropriate value is to translate $f(x)$ $\frac{b}{2}$ units to the right. Calculating

$$\begin{aligned} f\left(x - \frac{b}{2}\right) &= \left(x - \frac{b}{2}\right)^2 + b\left(x - \frac{b}{2}\right) + c \\ &= x^2 - bx + \frac{b^2}{4} + bx - b\frac{b}{2} + c \\ &= x^2 + \left(-\frac{b^2}{4} + c\right). \end{aligned}$$

Consequently, the roots of $f\left(x - \frac{b}{2}\right)$ are $\pm\sqrt{\frac{b^2}{4} - c}$. Using our translation, the roots of $f(x)$ are then

$$-\frac{b}{2} \pm \sqrt{\frac{b^2}{4} - c}$$

as desired.

How do we apply this to the cubic equation? We need to figure out how much to translate the function to get the type of graph we want. If we translate the function $f(x) = x^3 + bx^2 + cx + d$ by δ , we obtain the equation

$$\begin{aligned} f(x - \delta) &= (x - \delta)^3 + b(x - \delta)^2 + c(x - \delta) + d \\ &= x^3 - 3\delta x^2 + 3\delta^2 x - \delta^3 + b(x^2 - 2\delta x + \delta^2) + cx - c\delta + d \\ &= x^3 + (b - 3\delta)x^2 + (c + 3\delta^2 - 2b\delta)x + (d - \delta^3 + b\delta^2 - c\delta). \end{aligned}$$

To eliminate the x^2 term, we then must let $\delta = \frac{b}{3}$. Plugging this in above, we have

$$f\left(x - \frac{b}{3}\right) = x^3 + \left(c + 3\left(\frac{b}{3}\right)^2 - 2b\frac{b}{3}\right)x + \left(d - \left(\frac{b}{3}\right)^3 + b\left(\frac{b}{3}\right)^2 - c\frac{b}{3}\right).$$

Simplifying this equation we obtain:

$$f\left(x - \frac{b}{3}\right) = x^3 + (c - b^2/3)x + (d + 2b^3/27 - bc/3).$$

We can now find the roots of $f(x - \frac{b}{3})$ by setting $p = (c - b^2/3)$ and $q = bc/3 - 2b^3/27 - d$ to obtain a value X which when plugged in for x in $f(x - \frac{b}{3})$ will yield 0. Once you have this value X , then $X - \frac{b}{3}$ will be a root of the function $f(x)$.

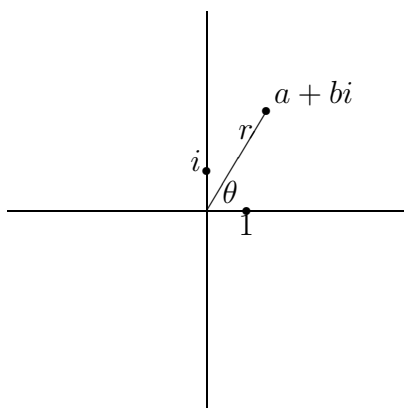
4.3 The Complex Plane

To see how to get all of the roots of a cubic, we need to analyze cube roots more and see what happens with the case of having a negative value inside of the square root. After all, we have found one root of a cubic polynomial, but we know in general that there should be three such roots. Where does our solution allow for this? It shows up in our solution when we move from u^3 to u . That is, we take a cube root, but there are really three different choices for the cube root of a number when we work in the complex numbers.

Before going into too much detail here, we need to examine the complex numbers more deeply, and we would like to attach some geometric meaning to them. In the last chapter, we noted that we could treat extension fields as vector spaces. In particular, thinking of the Complex numbers as a two dimensional vector space over the real numbers and taking $\{1, i\}$ as a basis for this vector space, the complex number $a + bi$ corresponds to the matrix

$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. We shall use this representation of the complex numbers together with polar coordinates to obtain our geometric realization.

Thinking of the complex numbers as a two dimensional vector space over the real numbers, we obtain a map between the complex numbers and the Cartesian plane, where the number $a+bi$ corresponds to the point (a, b) on the plane. As discussed in the high school curriculum and calculus, in addition to rectangular coordinates on the plane, we can use polar coordinates. In this case, the number $a+bi$ corresponds to the coordinate (r, θ) , where $a = r \cos \theta$ and $b = r \sin \theta$. The difficulty with this polar representation is that addition becomes quite difficult to define.



On the other hand, if we use the polar representation in our matrix form, then the matrix becomes $\begin{pmatrix} r \cos \theta & -r \sin \theta \\ r \sin \theta & r \cos \theta \end{pmatrix}$. This matrix, however, can be written as a product of two matrices

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

The two matrices here correspond to a dilation (or contraction) of the coordinate plane by a factor of r and a rotation of the coordinate plane by the angle θ . Thus, the complex number $a + bi$ can be thought of as representing a geometric action on the coordinate plane. Moreover, since dilations and contractions commute with the rotations, we can quickly see how to multiply two complex numbers that are given in polar coordinates. That is, the product of the complex numbers having polar coordinates (r, θ) and (s, ϕ) is the complex number having coordinates $(rs, \theta + \phi)$.

The piece of the puzzle of complex numbers comes from the work of Euler. In working with the exponential function and its Taylor series, Euler examined what happens when evaluating e^{ix} . Of course, the meaning of raising a number to a non-rational power is not easily described. When we first discuss raising a number to a positive integer power, we describe it as representing a repeated multiplication. We extend exponentiation to the rational numbers by taking roots. The next extension to irrational powers, on the other hand, **must** be done using limits in one way or another. Of course, in the high school curriculum, we frequently gloss over this fact, but we should remember that limits are there in the background. Even with limits, however, it is unclear what we should mean when raising a number to a complex power. The exponential function e^x being somewhat easier to deal with than other such functions as it has a nice Taylor's series for the function,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots$$

allows us a method to attack this question. Euler's solution was to simply use ix in place of x in the above equation, yielding:

$$e^{ix} = 1 + \frac{ix}{1!} + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \dots$$

Using that $i^2 = -1$, and bringing the terms together that involve i , we obtain

$$e^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots\right) + i \left(\frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots\right).$$

Examining these two terms more closely, however, we see that they correspond to the Taylor series for $\cos x$ and $\sin x$ respectively, giving Euler's great formula

$$e^{ix} = \cos x + i \sin x.$$

Thus, using the complex number $r \cos \theta + ir \sin \theta$ can be written as $re^{i\theta}$, and in this form, the multiplication rule $re^{i\theta} \cdot se^{i\phi} = rse^{i(\theta+\phi)}$ becomes easy to remember. Moreover, using this, taking roots becomes simplified. Namely, if you want the n th root of $re^{i\theta}$, you can simply raise this number to the $\frac{1}{n}$ power, yielding

$$\left(re^{i\theta}\right)^{\frac{1}{n}} = r^{\frac{1}{n}}e^{i\frac{\theta}{n}}.$$

The above is only one root among n , however. To get the other roots, we use the identity

$$e^{i\theta} = e^{i(\theta+2k\pi)},$$

where k is an arbitrary integer to arrive at the general solution

$$\left(re^{i\theta}\right)^{\frac{1}{n}} = r^{\frac{1}{n}}e^{i\frac{\theta+2k\pi}{n}}.$$

This gives us n different solutions for $k = 0, 1, \dots, n - 1$. At this point, we can now finish the solution of the cubic.

Given an arbitrary cubic equation $rx^3 + sx^2 + tx + u = 0$ with r, s, t , and u all real numbers and $r \neq 0$, we begin by dividing the equation through by r , to obtain an equation of the form $x^3 + bx^2 + cx + d = 0$ where $b = s/r$, $c = t/r$ and $d = u/r$. At this point, we let $x = x' - b/3$ to obtain the equation $(x')^3 + px = q$, where $p = (c - b^2/3)$ and $q = bc/3 - 2b^3/27 - d$ as we saw earlier. At this point, we can solve for x' to obtain that $x' = u - v$, where

$$u = \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} \quad \text{and}$$

$$v = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

While we have three choices for both u and v , we have one further constraint in that $3uv = p$. Thus, we choose our cube roots so that we have this requirement, which can be done simply by checking that the angles add up to either π or 0 depending on whether p is negative or positive. Then we have three choices for $x' = u - v$, and $x = x' - b/3$.

Aside

At this point, we have looked at number as ratio, as magnitude, and as being built up from roots. None of these really account for how we can think of negative numbers. In fact, in the history of European mathematics, negative numbers came after all of these understandings. Descartes, among others, considered negative numbers as fanciful constructs. Even today, one of the most difficult tasks of the middle school teacher is getting students

to understand why the product of two negative numbers should be a positive number. Sfarid argues that one cannot make this argument successfully without using our desire for the number system to satisfy the distributive law [14], or giving a new understanding of number. Complex numbers have a similar difficulty associated to them. What we have seen in this section, however, is that we can give these numbers a geometric meaning, if we consider them not as quantities that measure something, but rather as functions on the Euclidean plane (like matrices).

This understanding of number, opens up a whole new realm of mathematics. Numbers need no longer have a specific meaning. Instead, numbers are classified by the properties they have. This allows for one to consider more and more general sorts of “numbers,” and even to analyze what the properties themselves imply. This analysis of the properties devoid of the actual entities becomes the basis of abstract algebra, the study of groups, rings, fields, and integral domains. Indeed, in modern mathematics, the real number system is one of many different systems that extends the rational numbers.

End of Aside

4.4 Algebraic Numbers

We have seen that we can solve a cubic equation by using radical signs, and the operations of addition, subtraction, multiplication, and division on the coefficients. After the general cubic was solved, the next step in the mathematics of solving equations was to solve the general quartic or fourth degree equation by radicals. In current times, this is done most easily by first solving a cubic that arises from the original equation (much like we used a quadratic equation for our cubic solution) and then use this cubic solution to allow us to change the problem to an easier quartic equation that we could solve. The next step was to solve the quintic equation by radicals. Unfortunately, try as hard as they might, no one could find a solution by radicals for many years. Finally, it began to dawn on mathematicians that perhaps no solution could be found. This result was then first proved by Niels Abel in the 1800s, and then later Evariste Galois was able to create a theory, which allowed mathematicians to determine precisely when a polynomial could be solved by radicals. While Galois theory is beautiful and wonderful to study, it is beyond the scope of this class, as is Abel’s proof of the impossibility of solv-

ing the general quintic by radicals. (A great summary of this history can be found in [16].)

If we cannot solve every quintic by radicals, this means that there must be real numbers which cannot be found using the operations of addition, subtraction, multiplication, division, and taking n th roots on the rational numbers. Thus, yet again, we need to extend our definition of what a number should be. Our study of the complex numbers, however, has also given us a certain amount of freedom, as we might now realize that we can allow for numbers to be defined less concretely.

We say the real number α is *algebraic* if α is the root of some polynomial with rational coefficients. For example, $\sqrt[3]{3}$ is algebraic since it is the root of the polynomial $x^3 - 3$. If a number is not algebraic, then we say that it is *transcendental*. It is unclear whether any number can be transcendental at first glance. Certainly, most numbers that we might write down are algebraic. It turns out that both e and π are transcendental, although this is quite difficult to show. Our goal in this section will be to show that the algebraic numbers form a subfield of the real numbers, and then to show that one specific number is transcendental via a technique due to Liouville.

As we saw earlier, if α is a root of the polynomial $p(x)$, and $p(x)$ is irreducible over the rational numbers, then $Q[\alpha]$ is finite dimensional over Q . Now, suppose β is also the root of an irreducible polynomial with rational coefficients. Let $q(x)$ be a polynomial of minimal degree having β as a root. Again, we have that $[Q[\beta] : Q]$ is finite. Unfortunately, we do not yet have the matter in hand to show that $Q[\alpha, \beta]$ is finite dimensional over Q . However, we know that $g(x)$ is a polynomial over $Q[\alpha]$ since all of its coefficients lie in $Q \subseteq Q[\alpha]$. Thus, a polynomial of $Q[\alpha]$ of minimal degree having β as a root, must have degree smaller than the degree of $g(x)$. Consequently, $[Q[\alpha, \beta], Q[\alpha]] \leq [Q[\beta], Q]$, and $Q[\alpha, \beta]$ is finite dimensional over Q since it is the product of $\dim_{Q[\alpha]}(Q[\alpha, \beta])$ and $\dim_Q(Q[\alpha])$, both of which are finite.

At this point, let us set $E = Q[\alpha, \beta]$. As E is a field, $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, and α/β are all elements of E . Thus, if we can show that any element of E must be a root of some polynomial with rational coefficients, we would have that this is true for $\alpha\beta$, $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, and α/β . But this last is precisely what we need to see that the algebraic numbers form a field! Thus, we have reduced that problem of showing the algebraic numbers are a field, to showing that any element of E must be the root of a (non-zero) polynomial with rational coefficients.

Let $\gamma \in E$, suppose $\dim_Q(E) = k$, and consider the set $\{1, \gamma, \gamma^2, \gamma^3, \dots, \gamma^k\}$.

Since this set has $k + 1$ elements, it follows that it is linearly dependent over Q . That is, there exist elements $a_0, a_1, \dots, a_k \in Q$ not all equal to 0, such that

$$\sum_{i=0}^k a_i \gamma^i = 0.$$

This, however, implies that γ is a root of the polynomial

$$\sum_{i=0}^k a_i x^i.$$

Thus, we have shown:

Theorem 4.1 *The algebraic numbers form a field.*

4.5 Transcendental Numbers

The question now becomes: Is every real number algebraic? Euler gave the first definition of transcendental numbers, numbers that are not the root of any nonzero polynomial with rational coefficients. Euler, however, was unable to prove that any number was transcendental, although there is good evidence that he believed that both e and π were transcendental. The proof, however, of the existence of a transcendental number had to wait about 50 years, until 1844 when Liouville proved that $\sum_{n=1}^{\infty} 10^{-n!}$ was transcendental (along with many other similar numbers). While this result proved the existence of transcendental numbers, it did not show that any particularly interesting number (at the time) was transcendental. In 1873, however, Hermite proved that e was transcendental. In 1874, Cantor produced a startling result showing that transcendental numbers existed, although it did not produce any. More curious, Cantor showed that most real numbers are in fact transcendental, even though at the time, only few numbers could be written down and proved to be transcendental. By 1883, Lindemann showed that π was transcendental. Lindemann's work, while settling the question of the squaring of the circle, however, is not considered quite as grand as Hermite's, since Lindemann essentially used the same techniques as Hermite in a little bit more generality. In 1900, Hilbert presented a list of 23 problems for mathematics to answer in the new century. The seventh problem of Hilbert was to determine whether or not $2^{\sqrt{2}}$ was transcendental. The problem was

answered in 1934 when A.O. Gelfond proved the result (also done independently in 1935 by T. Schneider) that if α and β are algebraic numbers with $\alpha \neq 0$, $\alpha \neq 1$, and β not a real rational number, then any value of α^β is transcendental. (See [7] pp.134-150 for a historical treatment and proof of this remarkable result, while the original works by Gelfond and Schneider can be found in [3] and [12], although more readable expositions of these results can be found in [4].)

How can one show that a number α is transcendental? To do so, requires proving somehow that no non-zero polynomial with rational coefficients can have α as a root. We can make things a little easier by reducing this problem to the case where the polynomials are only allowed to have integer coefficients, since we can multiply through by a common denominator of the coefficients. As with the proofs of irrationality, we need to use proof by contradiction in this case. That is, we will need to assume that our candidate α is the root of a polynomial with integer coefficients, and then obtain a contradiction. Liouville was able to use this method and calculus to show that the number $\sum_{n=1}^{\infty} 10^{-n!}$ was transcendental. We now prove this.

Theorem 4.2 *The real number $\alpha = \sum_{n=1}^{\infty} 10^{-n!}$ is transcendental.*

Proof: Suppose that $p(x)$ is a polynomial with integer coefficients such that $p(\alpha) = 0$. We shall aim for a contradiction by showing that if α is the root of this polynomial, then there are only finitely many rational numbers $\frac{a}{b}$ such that $|\alpha - \frac{a}{b}| < b^{-n-1}$ where n is the degree of $p(x)$. Using this together with $|\alpha - \sum_{n=1}^k 10^{-n!}| < 10^{-(k+1)!}$, will give us the desired contradiction since $\sum_{n=1}^k 10^{-n!} = \frac{a}{b}$ where $b = 10^{-k!}$.

Now, suppose that $a, b \in \mathbb{Z}$ are such that $p(\frac{a}{b}) \neq 0$ (note that there are only finitely many rational numbers that do not satisfy this property). Since the coefficients of $p(x)$ are all integers, it must be the case that $p(\frac{a}{b}) = \frac{A}{b^n}$ for some integer $A \neq 0$, where $n = \deg(p(x))$ (show this!).

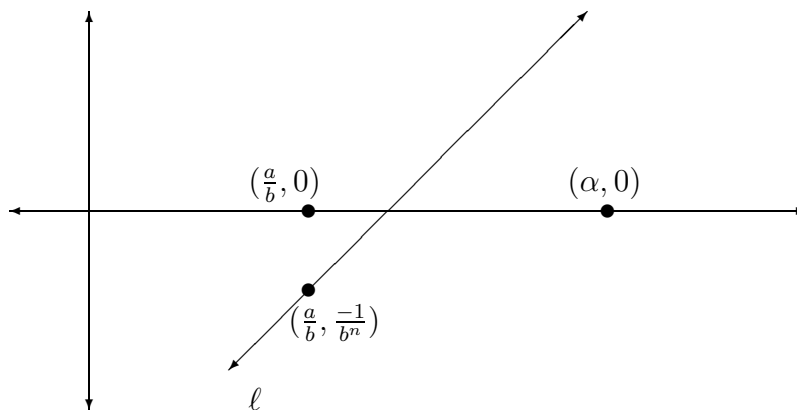
The second step in the proof is to get some control on the values that $p(x)$ can take on nearby α . Since $[\alpha - 1, \alpha + 1]$ is a closed interval and $p(x)$ is a polynomial, it follows that $|p'(x)| \leq m$ for some integer m on the interval $[\alpha - 1, \alpha + 1]$.

At this point, we suppose $\frac{a}{b} \neq \alpha$ is a rational number in the interval $[\alpha - 1, \alpha + 1]$. For the time being, we shall assume that $\frac{a}{b} < \alpha$ and that $p(\frac{a}{b}) < 0$, and is therefore less than or equal to $\frac{1}{b^n}$. As $p'(x) < m$ on $[c, d]$, if

follows that the graph of $p(x)$ lies below the graph of the line ℓ given by the equation

$$\left(y + \frac{1}{b^n}\right) = m\left(x - \frac{a}{b}\right).$$

More precisely, $p(x) \leq m\left(x - \frac{a}{b}\right) - \frac{1}{b^n}$.



This can be established by using the mean value theorem from calculus, which implies that if the inequality is not satisfied, then $p'(x) > m$ for some x in the interval. Alternatively (and probably more intuitively, one can apply the racetrack principle [18]). (The technical proof with the mean value theorem is that $p(x)$ is continuous on the interval $[\frac{a}{b}, x]$ with the derivative continuous on $(\frac{a}{b}, x)$, so that the mean value theorem implies for some $c \in (\frac{a}{b}, x)$, we have

$$f'(c) = \frac{p(x) - p(a/b)}{x - \frac{a}{b}}.$$

Since $f'(c) \leq m$, it follows that

$$p(x) \leq mx - m\frac{a}{b} + p(a/b) \leq mx - m\frac{a}{b} - \frac{1}{b^n},$$

where the last term follows because we have assumed that $p(a/b) < 0$ so that $p(a/b) < \frac{1}{b^n}$ from above.)

Thus, the value of α is greater than the x -coordinate of the intersection point of ℓ and the x -axis. We can calculate this coordinate, however, and it is $\delta = \frac{a}{b} + \frac{1}{mb^n}$. Consequently,

$$\left| \alpha - \frac{a}{b} \right| \geq \frac{1}{mb^n}.$$

If $b > 2m$, however, we then have

$$\left| \alpha - \frac{a}{b} \right| \geq \frac{2}{b^{n+1}}.$$

Choosing $t > 2m$ sufficiently large, it follows that for any $b > t$, if $\left| \alpha - \frac{a}{b} \right| \leq \frac{2}{b^{n+1}}$ it must be the case that $\frac{a}{b} \in [c, d]$. However, we also just showed that if $\frac{a}{b} \neq \alpha$ then $\left| \alpha - \frac{a}{b} \right| > \frac{2}{b^{n+1}}$. Consequently, there can only be finitely many rational numbers $\frac{a}{b}$ having the property

$$\left| \alpha - \frac{a}{b} \right| < \frac{2}{b^{n+1}}.$$

(Note that if we had assumed that $p(a/b) > 0$ and/or $a/b > \alpha$, we would merely have changed which line we were looking at, and we still would obtain the same final result.) At this point, we are nearly done with the proof that $\alpha = \sum_{n=1}^{\infty} 10^{-n!}$ is transcendental. Let $q_k = \sum_{n=1}^k 10^{-k!}$. Then

$$|\alpha - q_k| = \sum_{n=k+1}^{\infty} 10^{-n!} \leq \frac{2}{10^{(k+1)!}}.$$

However, $q_k = \frac{a_k}{b_k} = \frac{a_k}{10^{k!}}$ for some integer a_k , and $b_k = (10^{k!})^{n+1} = 10^{(n+1)(k!)}$. If $k > n$, however, we then have the difference

$$\begin{aligned} |\alpha - q_k| &< \frac{2}{10^{(k+1)!}} \\ &= \frac{2}{10^{(k+1)(k!)}} \\ &< \frac{2}{10^{(n+1)(k!)}} \\ &= \frac{2}{b_k^{n+1}}. \end{aligned}$$

Since q_k has the property for all $k > n$ that $|\alpha - q_k| < b_k^{n+1}$, it follows that α cannot be algebraic.

Q.E.D.

Most analytic number theory texts prove the general statement of Liouville's theorem, which can be read out of our proof. Namely that

Theorem 4.3 *If α is a real algebraic number, then there exists only finitely many rational numbers $q = \frac{a}{b}$ (where a and b are integers with $b > 0$) such that $|\alpha - q| < \frac{2}{b^{n+1}}$.*

The choice for this presentation was to concentrate on one specific number.

Aside

Why go through this proof at all? I have two reasons. The first is that we frequently tell our students that e and π are transcendental numbers, without truly understanding what this means. Unfortunately, the proofs for the transcendence of e and π are both substantially more difficult than the proof that we gave in Chapter 2 that π was irrational, and consequently, it is really beyond the high school curriculum. On the other hand, the only material really needed for the proof of Liouville's theorem are the concept of slope or rate of change and the understanding of what value polynomials with integer coefficients can take on rational numbers. While some might take issue, pointing out that we used derivatives and continuity in our proof, both of these concepts can be understood at a lower level by drawing graphical representations of polynomials, while leaving the capital P Proof for later courses. Moreover, one can turn the Mean Value Theorem with the race track lemma (as put in [18]), that can be put simply as: if two race tracks are going around a track, and the first car is always going faster than the second car, then the first car is always ahead of the second car. This readily believed lemma on race cars works with our polynomials to establish that the polynomial must lie between the two lines.

Thus, from this one can get a better understanding of some of the properties distinguishing algebraic numbers from transcendental numbers from this proof. In particular, the idea that there is a limit to how well one can approximate algebraic numbers, whereas at least some transcendental numbers can be approximated remarkably well. You should note in fact, the similarity

that this argument has with several of our proofs of irrationality. The one difference is that the level of approximation required is much tighter in the transcendence proof than it was in the irrationality proof.

At this point, it makes sense to emphasize the role of proof in mathematics. The standard argument is that proof allows us to know something with a certainty. While this certainty is important in many cases, what is probably more important is the role of proof in mathematical problem solving. When solving a problem in mathematics, proof can play a very important part. Schoenfeld says [13]

For the mathematician, dependence on argumentation as a form of discovery is learned behavior, a function of experience. This perspective is not “natural”; few B.A.s in mathematics possess it. Those who become mathematicians generally develop this perspective in graduate school, during their apprenticeship to the discipline. At first, “proof” is mandatory, an accepted standard. As one becomes acculturated to mathematics, it becomes natural to work in such terms. “Prove it to me” comes to mean “explain to me why it is true,” and argumentation becomes a form of explanation, a means of conveying understanding. As the mathematician begins to work on new problems (perhaps on a dissertation), this progression continues. Mathematical argument becomes a way of convincing oneself that something ought to be true. Even unsuccessful attempts turn out to be valuable, because consistent failures point to weak spots in one’s understanding. After numerous attempts to demonstrate a particular result, one can see a pattern in the failures and decide the result may not be true. To see if it is false, one may try to construct some examples that exploit the weak spots. If none of the examples one tries demonstrate that the result is false, one may again begin to believe the result is true - and a pattern in the failed examples may suggest the information that was missing from the original attempts. Thus mathematical argument becomes a tool in the dialectic between what the mathematician *suspects* to be true and what the mathematician *knows* to be true. In short, deduction becomes a tool of discovery. (pp.172-173)

An important point here is the idea of mathematical argument as a tool for understanding and learning. For example, the proof that the product of two

odd integers is odd, can be used several different ways in a classroom. If the teacher presents it, then it becomes an exercise in writing a proof. In this fashion, however, students will disengage from the process, or learn it by rote without understanding. A second method can be for the students to complete the proof themselves, may well feel like busy work to the students. A third method is to have the students analyze the proof and see what they can say if odd is replaced by divisible by three. The natural proof (before Euclid's Lemma is introduced to students) for the odd multiples is to write down two arbitrary odd numbers as $2k + 1$ and $2l + 1$ and to multiply them out. When trying to see how this works for the case of 3, the students can discover how to break the problem into separate cases, to arrive at arithmetic modulo 3. Similarly, when discussing geometric constructions, it is the knowledge of what theorems are true *and how to apply them* that is crucial in developing geometrical constructions.

For the above reasons, one of the important tools to learn in any mathematics class is the role of proof in discovery. Reasoning plays a major role in the NCTM standards. One major understanding is the varying levels of proof (from heuristic and intuitive argumentation to formal) and their role in problem solving. One possible interpretation (mine) for how teachers can help students develop mathematical reasoning and problem solving skills is by varying the levels as one moves through the curriculum.

End of Aside

Chapter 5

Dedekind Cuts

5.1 Axioms for the Real Numbers

We are about to embark on the explicit construction of the real numbers from the rational numbers. Before doing so, however, you should work through the following questions. The purpose of these questions are to encourage you to think about what properties we would like the real numbers to satisfy and to work from a minimum set of axioms to define the real numbers.

To begin with, consider the following:

The guiding principle in defining the real numbers is the number line. What are all the properties that the set of real numbers should have? You should come up with at least twelve properties that are expected of the set (Hint: Think about the entire field).

Below we list the actual defining axioms for the real numbers. You may well not have gotten all of them. In fact, it is fairly typical that students will miss out on all of the axioms involving the order relationship $<$. One of the reasons for this is that we tend to take the idea of order on the number line for granted. Thus we tend to take it as a given that the order relation exists. Moreover, we also often take it for granted that this relationship

The real numbers R are a set together with the two binary operations $+$: $R \times R \rightarrow R$, \cdot : $R \times R \rightarrow R$, and the binary relation $<$ on R under the following thirteen axioms:

1. The operations $+$ and \cdot are associative.
2. The operations $+$ and \cdot are commutative.
3. The operation \cdot is distributive over $+$. That is, for all $a, b, c \in R$, $a \cdot (b + c) = a \cdot b + a \cdot c$.
4. There exist at least two distinct elements in R .
5. There exists an element $0 \in R$ such that for all $a \in R$, $a + 0 = a$.
6. For each element $a \in R$, there exists an element $b \in R$ such that $a + b = 0$. We call this element $-a$.
7. There exists an element $1 \in R$ such that for all $a \in R$, $a \cdot 1 = a$.
8. For each element $a \in R$, with $a \neq 0$ there exists an element $b \in R$ such that $a \cdot b = 1$. We call this element a^{-1} or $\frac{1}{a}$.
9. The relation $<$ is transitive on R .
10. (The Trichotomy Law). For every $a, b \in R$, exactly one of $a < b$, $a = b$, and $a > b$ is true.
11. For all $a, b, c \in R$, if $a < b$ then $a + c < b + c$.
12. For all $a, b, c \in R$, if $a < b$ and $c > 0$, then $a \cdot c < b \cdot c$.
13. (The Greatest Lower Bound Axiom) If $S \subset R$ is a non-empty subset, and there exists a lower bound $b \in R$ such that for all $a \in S$ we have $b \leq a$, then there exists a greatest lower bound c for S . That is, there exists $c \in R$ such that $c < a$ for all $a \in S$ and moreover, if $b \in R$ also satisfies the condition $b < a$ for all $a \in S$, then $c \leq b$.

Using the axioms above, let us move on to prove some things that we expect about the real numbers.

1. If $x < y$, then $-y < -x$. Hint: Assume $x < y$ and add the same thing to both sides of the inequality in order to establish that $-y < -x$.
2. $0 < 1$. Although this statement seems obvious, you need to establish this result using the pink sheets axioms. What are the three possible ways of relating 0 and 1? Use the axioms to establish that what you know is impossible is truly impossible.
3. If $0 < x < y$, then $0 < \frac{1}{x} < \frac{1}{y}$.
 - (a) Assume $0 < x < y$. What are the possibilities for the relationship between 0 and $\frac{1}{x}$?
 - (b) Argue why, from the set of axioms, two of these possibilities cannot be true.
 - (c) What does this mean about $\frac{1}{y}$?
 - (d) Using $0 < x < y$, multiply each element of the compound inequality by the same factor in order to establish $0 < \frac{1}{y} < \frac{1}{x}$. Explain why this is legal.
4. If $x < y$ and $z < 0$, then $yz < xz$. Hint: Assume $x < y$ and $z < 0$. Use axiom 11 and a previous result to establish $yz < xz$.
5. Prove the Theorem:

Theorem 5.1 *If S is a non-empty set of real numbers that is bounded from above, then S has a least upper bound.*

- (a) Argue why if M is an upper bound for S , then $M \geq s$ for all $s \in S$.
- (b) Let $T = \{t \mid t = -s, \text{ for some } s \in S\}$. Argue why $-M$ is a lower bound for T .
- (c) Explain why if K is a lower bound for T , then $-K$ is an upper bound for S .
- (d) Argue why T must have a greatest lower bound, call it B .

- (e) Argue why $-B$ must be an upper bound for S .
- (f) If C is an upper bound for S , then what is the relationship between $-C$ and B as well as $-C$ with the set T ?
- (g) What is the relationship between C and $-B$? How does this imply $-B$ is the least upper bound for S ?

6. Prove the following theorem:

Theorem 5.2 *Let x be any real number. Then there is an integer $n \leq x < n + 1$.*

In order to accomplish this, we need to define the set $A = \{n \mid n \in \mathbb{Z} \text{ and } n \leq x\}$.

- (a) If A is non-empty, argue why A is bounded from above and has a least upper bound (call it n_0).
- (b) Explain why if you subtract 1 from n_0 , the resultant value will not be an upper bound of A .
- (c) Discuss why there exists an $m \in A$ such that $n_0 - 1 < m \leq n_0$.
- (d) Explain why $m + 1 \notin A$ and $x < m + 1$.
- (e) How does this establish, for the case that A is non-empty, the result that we can find an integer n such that $n \leq x < n + 1$.
- (f) To show that A is not empty, consider the set $B = \{b \mid b > x \text{ and } b \in \mathbb{Z}\}$, and show that it must have a least upper bound called b_0 . Argue that there exists a $m \in B$ such that $b_0 \leq m < b_0 + 1$ so that $m - 1 \in A$.

7. Prove the following theorem

Theorem 5.3 *Between any two real numbers is both a rational number and an irrational number.*

- (a) Let x and y be real numbers with $x < y$. Why is there an integer N such that $0 < \frac{2}{y-x} < N$?
- (b) Argue why there exists an integer $n \leq Nx < n + 1$.
- (c) Argue why $n + 2 < Ny$.
- (d) Explain why $\frac{n+1}{N}$ and $\frac{n+2}{N}$ are rational numbers between x and y .
- (e) Explain why $\frac{n+\sqrt{2}}{N}$ is an irrational number between x and y .

5.2 Dedekind Cuts

Descartes' Geometry is an important advance in mathematics, because it gives the first instance of a mathematician writing equations for geometric figures (although, the equation of a line occurs only once within the text [15]) and it gives the first idea of calculating the normal to a curve. Moving forward in time, Leibniz and Newton develop calculus using the idea of the coordinate axes and rates of change, and as we have seen earlier, questions about algebraic and transcendental numbers begin to be raised by Euler in the 1700s. Cauchy defines and "proves" that Cauchy sequences converge in the early 1800s, but curiously, it is not until the 1890s that mathematicians realized that they did not have a proper definition of the real numbers. In some ways, this isn't so surprising. The move to axiomatize all mathematical systems is a product of the late 1800s and the discovery of non-Euclidean geometries, and the real numbers seen as points on the number line are well defined by intuitive standards. Moreover, calculus and analysis were extremely successful without worrying about foundational issues. On the other hand, as we shall see later, even such great mathematicians as Cantor, were prone to make subtle errors when dealing with the real numbers.

In the 1800s, Richard Dedekind taught a calculus course in Germany. He decided that he wanted to start from the basic definitions and put calculus on a firm foundation. As he began to design the course, he ran into trouble proving the intermediate value theorem, one of the important foundational theorems of the calculus. After struggling with this, he realized that the problem was that he had no strict definition of the real numbers to work from.

Theorem 5.4 (The Intermediate Value Theorem) *If $f(x)$ is a continuous function on the closed interval $[a, b]$ with $f(a) < 0$ and $f(b) > 0$, then there exists $c \in (a, b)$ such that $f(c) = 0$.*

The proof of this theorem uses the greatest lower bound axiom. Namely that every non-empty set of real numbers that has a lower bound has a greatest lower bound. To prove the theorem, one considers the set $C = \{x \in [a, b] \mid f(x) \geq 0\}$. The set C is non-empty since $b \in C$, and the set has a lower bound since $a < x$ for all $x \in C$. Thus C has a greatest lower bound, which we shall call c . At this point, all that remains is to prove that $f(c) = 0$. To see this, we use the continuity of f . By continuity and the knowledge that $f(b) > 0$ and $f(a) < 0$, it follows that $c \in (a, b)$. Now, since for all $d \notin C$,

we know that $f(d) \leq 0$, we have that $f(c) \leq 0$ as $f(x)$ is continuous. On the other hand, for all $\delta > 0$, there exist a $y \in C$ such that $y < c + \delta$ since c is a greatest lower bound. However, then $f(y) \geq 0$. Thus, continuity again implies that $f(c) \geq 0$, leaving us with $f(c) = 0$.

This is the modern proof of the theorem, where we take one of the completeness axioms as part of the axiomatic definition of the real numbers. Consider Dedekind's problem, however. The real numbers are not axiomatically defined, but rather they are intuitively defined by the number line. The greatest lower bound axiom is only true intuitively in this situation, and can hardly be called on in a late 19th century perspective. Thus, Dedekind needed to construct the real numbers in some way. His insight was to use the intuition of the number line in his construction. That is, he reasoned that every point on the number line breaks the real line into two separate pieces, and he wanted to use the pieces to identify the numbers. The down side of this attempt is that the pieces consist intuitively of real numbers, and you cannot use the real numbers in defining what the real numbers are! On the other hand, the rational numbers were well-defined at this time, so Dedekind reasoned that to determine the two pieces of the number line, it was sufficient to consider only the rational numbers in each piece. From this intuition, Dedekind gave the following definition:

Definition: A *Dedekind cut* is a pair (A, B) of non-empty sets such that

1. $A \cup B = Q$,
2. $A \cap B = \emptyset$, and
3. If $a \in A$ and $b \in B$, then $a < b$.

Dedekind's goal was to define the real numbers as the set of Dedekind cuts. However, one needs to be careful about this. If the intuition is to think about how the line breaks into pieces, the point at which you break the line needs to end up on one side or the other. If the point that you break the number line at is not rational, this isn't a problem since the pieces now consist only of the rational numbers. However, if you break the line at a rational point, then you do have a problem, and a choice has to be made about which one of the two sets to put it in. We will take Dedekind's point of view, which is different from most modern authors, in that we will consider each of these to be a Dedekind cut. That is, given a rational number q , there

are two cuts associated to q , namely the cut $(A_{\underline{q}}, B_{\underline{q}})$ given by

$$\begin{aligned} A_{\underline{q}} &= \{a \in Q \mid a \leq q\} \\ B_{\underline{q}} &= \{b \in Q \mid b > q\}, \end{aligned}$$

and the cut $(A_{\overline{q}}, B_{\overline{q}})$ given by

$$\begin{aligned} A_{\overline{q}} &= \{a \in Q \mid a < q\} \\ B_{\overline{q}} &= \{b \in Q \mid b \geq q\}. \end{aligned}$$

In the modern treatment, a Dedekind is defined slightly differently. In particular, a cut is usually a single set corresponding to only one of our two sets.

We define an equivalence relation on Dedekind cuts so as to ensure that the two cuts associated to the same rational number are equivalent. To this end, we say that the cuts (A, B) and (C, D) are *equivalent* if $(A \cup C) \setminus (A \cap C)$ is a set consisting of at most one element. That is, the sets A and B differ by at most one element.

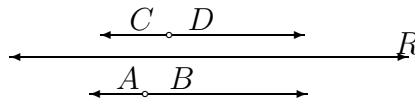
We now define the real numbers as the set of equivalence classes of Dedekind cuts. At this point, we need to define the operations and order relation. One can make most of these definitions easier by assuming that if (A, B) is a rational cut, then the corresponding rational number is an element of B . While we will not do this in general, we will state these definitions in parentheses after our definitions. The reason for this change is that we want to be able to work with either element of the equivalence class. Basically, you either have to work with a particular element of each equivalence class or worry about proving that the operations we shall define are well defined.

We define the order relation $<$ on R as follows: Given two non-equivalent Dedekind cuts (A, B) and (C, D) , we say that $(A, B) < (C, D)$ if and only if $A \subset C$. (The condition is not simplified by the assumption that if the cut is rational then the rational point lies on the left hand side.) This condition implies in our intuitive understanding of Dedekind cuts that the breaking point of the number line determined by (A, B) lies to the left of the breaking point of the number line determined by (C, D) . We need to check that this definition is well defined. Namely, that if we choose cuts equivalent to (A, B) and (C, D) , that we will not change the definition of the relation $<$.

Suppose now that $(A, B) < (C, D)$ and $(A', B') \sim (A, B)$ and $(C', D') \sim (C, D)$. By our assumptions, there exists $c \in C$ such that $c \notin A$. Since

$(A, B) \not\sim (C, D)$, there are at least two elements c, c' of C that are not in A . By the third condition for Dedekind cuts, $\frac{c+c'}{2}$ and $\frac{2c+c'}{3}$ are also not elements of A . As $(A', B') \sim (A, B)$, A' contains at most one point that is not contained in A . Similarly, C and C' differ by at most one element. As C contains at least four elements not in A , C' must contain at least two elements not in A' , implying that $A' \subset C'$ by condition 3. Since \sim is an equivalence relation, we have that $(A', B') \not\sim (C', D')$, so that $(A', B') < (C', D')$ as we desired, proving that the relation $<$ is well defined.

Pictorially, we have the following:



While here we have concentrated on the left hand sets, we could similarly have concentrated on the right hand sets. The following proposition establishes this:

Proposition 5.5 *If $(A, B) \not\sim (C, D)$ are two Dedekind cuts, then $(A, B) < (C, D)$ if and only if $D \subset B$.*

Proof: We first prove the forward direction. Suppose $(A, B) < (C, D)$. Then $A \subset C$, implying there exists $c \in C$ such that $c \notin A$. Since $A \cup B = \mathbb{Q}$, it follows that $c \in B$. By the third condition for Dedekind cuts it follows that $a < c$ for all $a \in A$. If $d \in D$, then by the third condition for Dedekind cuts, it follows that $c < d$. Consequently, if $a \in A$, then $a < d$. In turn, this implies that $d \in B$. Consequently, $D \subseteq B$. To see that $D \neq B$, it remains to note that $c \notin D$.

Conversely, if $(A, B) \not\sim (C, D)$ and $D \subset B$, then there exists an element $b \in B$ such that $b \notin D$. By the first condition for Dedekind cuts, it follows that $b \in C$. By the third condition, we then have that for all $d \in D$, $b < d$. If $a \in A$, then it follows that $a < b$. Therefore, $a < d$ for all $d \in D$. Consequently, $a \notin D$, so that $a \in C$. Thus $A \subseteq C$, and the containment is strict since $b \notin A$.

Q.E.D.

Recall that (A_0, B_0) is one of the cuts associated to the rational number 0. To simplify our notation, we shall write $(C, D) > 0$ to denote that $(C, D) >$

(A_0, B_0) . In a similar way, we will write $(C, D) > q$ to denote that (C, D) is greater than either cut associated to q .

At this point, we need to define the binary operation of addition.

Definition: For Dedekind cuts (A, B) and (C, D) we define $(A, B) + (C, D)$ to be the cut (E, F) where

$$F = \{b + d \mid b \in B, d \in D\}$$

and $E = Q \setminus F$. (If we insist that rational cuts contain the rational point on the right, then we can write $E = \{a + c \mid a \in A, c \in C\}$. Again, we should check that this definition is well defined, but this is left to the reader.

Given that this definition is well defined, it is a straightforward check to see that addition is commutative and associative, and thus we have that the first two axioms are satisfied for addition.

To define multiplication of Dedekind cuts is unfortunately harder. In particular, the difficulty of multiplying negative numbers comes back to haunt us.

Definition: Given two Dedekind cuts $(A, B) > 0$ and $(C, D) > 0$, the *product* of these cuts is defined to be the cut (E, F) where

$$F = \{bd \mid b \in B, d \in D\}$$

and $E = Q \setminus F$.

Before moving on to define the product of negative cuts, let us first show that this definition is well defined. Suppose that $(A', B') \sim (A, B)$ and $(C', D') \sim (C, D)$, and let $F' = \{bd \mid b \in B', d \in D'\}$. Without loss of generality, assume that $D \subset D'$, and the containment is proper. Then it follows that $D' = D \cup \{q\}$ where q is a minimal element of D' . We claim that $F' \neq F$ if and only if B is a proper subset of B' , and in this case F' and F differ by at most one element. Suppose first that B' does not contain a minimal element. Then if $x = b'd'$ for some $b' \in B'$ and some $d' \in D'$, as B' does not contain a minimal element (implying that $B' \subsetneq B$, there exists a rational number $\tilde{b} \in B'$ such that $\tilde{b} < b'$. Consequently $\frac{x}{\tilde{b}} > d'$, and $\frac{x}{\tilde{b}}$ is therefore an element of D' and also of D . This implies however that $x = \tilde{b}d \in F$, so that $F' \subseteq F$. Next, suppose B' contains a least element r . Let $x \in F'$. By the previous argument, $x \in F$ if $x = bq$ for some $b \in B'$ with $b \neq r$. Similarly, the argument above would imply $x \in F$ if $x = rd$ for some $d \in D$ with $d \neq q$. Consequently, $x \in F$ but $x \notin F'$ implies $x = rq$, and F' contains at most one element more than F , and only in the case where B' and

D' both contain a minimal element but one of B or D does not. Reversing the roles of F and F' above yields that F and F' differ by at most one element, and thus $(E, F) \sim (E', F')$. This argument is tedious; it does however bring light to why the modern treatment simply insists on defining cuts in only one direction. Namely, by doing so, we can avoid both equivalence classes and the trivialities of proving the operations are well defined.

Actually, we still need to check that the product of two positive cuts is, in fact, a cut.

Proposition 5.6 *Let $(A, B) > 0$ and $(C, D) > 0$ be two Dedekind cuts, and let (E, F) be their product. Then (E, F) is also a Dedekind cut.*

Proof: Since B and D are both non-empty (as (A, B) and (C, D) are cuts, it follows that there exists $b \in B$ and $d \in D$ so that $bd \in F$, and hence F is non-empty. Since B and D consist only of positive numbers, so does F , so that $0 \in E$ and hence E is non-empty. The first two conditions for cuts are satisfied since $F \subset Q$ follows from the closure of the rational numbers under multiplication, and then defining $E = Q \setminus F$ forces $E \cup F = Q$ and $E \cap F = \emptyset$. Consequently, it remains to show the third condition. To see this, we shall first show that $x \in F$ and $y \in Q$ with $y > x$ implies $y \in F$. As $x \in F$, $x = bd$ for some $b \in B$ and $d \in D$. Since $y > x$ and $b > 0$, $y/b > x/b = d$. Consequently, y/b is a rational number, and by condition 3 for the Dedekind cut (C, D) implies that $y/b \in D$. Thus $y = b \cdot (y/b)$ is an element of F . Now suppose $e \in E$ and $f \in F$. If $e \geq f$, the above would imply that $e \in F$, contradicting that $E \cap F = \emptyset$. Thus $e < f$ as desired, and (E, F) is a Dedekind cut.

Q.E.D.

Before giving the general definition of multiplication, we next need to define the negative of a cut (A, B) . Given a cut (A, B) , let $-(A, B)$ denote the cut $(-B, -A)$ where

$$\begin{aligned} -A &= \{-a \mid a \in A\} \\ -B &= \{-b \mid b \in B\} \end{aligned}$$

To see that $(-B, -A)$ is a cut is straightforward as all but condition 3 follow immediately from the same conditions for (A, B) . The third condition requires a little more work. Suppose $x \in -B$ and $y \in -A$. Then $x = -a$ for

some $a \in A$ and $y = -b$ for some $b \in B$. By condition 3 for (A, B) , we have that $a < b$. Thus, using the rules of inequality for the rational numbers, we have $-b < -a$, and hence $x < y$ as desired. Therefore $-(A, B)$ is a Dedekind cut. To check that this is well-defined, suppose that $(A, B) \sim (C, D)$. Then A and C differ by at most one element. From this it follows that $-A$ and $-C$ differ by at most one element, implying that $(-B, -A) \sim (-D, -C)$.

Proposition 5.7 *If (A, B) is a Dedekind cut with $(A, B) < 0$, then $-(A, B) > 0$.*

Proof: From the above, we know that $-(A, B)$ is a Dedekind cut. Thus, it suffices to show that $-(A, B) > 0$. By the definitions, this requires us to show that $A_0 \subset -B$, and $-(A, B) \not\sim (A_0, B_0)$. Recall that this means that we need to show that if $x < 0$ then $x \in -B$, and that there exists some $y > 0$ such that $y \in -B$, as this would mean that $-B$ contains at least two points not in A_0 . But, $x < 0$ implies that $-x > 0$. As $(A, B) < 0$, we know that $B_0 \subset B$ by proposition 5.5, so that $-x \in B_0$ implies $-x \in B$. Thus $x \in -B$ as desired. Moreover, $(A, B) > (A_0, B_0)$ also implies that B contains at least two elements not in B_0 . I.e., B contains some negative rational number $-y$. Thus $y \in -B$, and $y > 0$. Consequently $-(A, B) > 0$.
Q.E.D.

We are now ready to define multiplication of two arbitrary Dedekind cuts. Let (A, B) and (C, D) be two cuts. Then

$$(A, B) \cdot (C, D) = \begin{cases} (A, B) \cdot (C, D) & \text{if } (A, B) > 0 \text{ and } (C, D) > 0 \\ -(-(A, B) \cdot (C, D)) & \text{if } (A, B) < 0 \text{ and } (C, D) > 0 \\ -((A, B) \cdot -(C, D)) & \text{if } (A, B) > 0 \text{ and } (C, D) < 0 \\ -(A, B) \cdot -(C, D) & \text{if } (A, B) < 0 \text{ and } (C, D) < 0. \end{cases}$$

That this is a well-defined definition follows as multiplications of positive cuts is well-defined, as is the definition of $-(A, B)$.

At this point, we can prove the associative, commutative, and distributive laws for multiplication and addition. The proofs, however, are tedious and not terribly informative. We shall prove the commutative law for addition, and leave it to the reader to check the others.

Proposition 5.8 *If (A, B) and (C, D) are any two Dedekind cuts, then $(A, B) + (C, D) = (C, D) + (A, B)$.*

Proof: Let $(A, B) + (C, D) = (E, F)$ and $(C, D) + (A, B) = (G, H)$. Then

$$F = \{b + d \mid b \in B, d \in D\} \tag{5.1}$$

$$= \{d + b \mid b \in B, d \in D\} \quad \text{by the commutative law for addition of rational numbers} \tag{5.2}$$

$$= H. \tag{5.3}$$

Thus $(E, F) = (G, H)$.

Q.E.D. This proof is similar to all of the proofs, in that they use the same property for the rational numbers to establish it for the real numbers (as Dedekind cuts). Note that this is similar to how we learn about the number systems. We first learn the rules for the natural numbers, and having them for the natural numbers, we extend the rule to the integers, having the rule for the integers, we extend it to the rational numbers. Thus, extending the rules to the real numbers from the rule for the rational numbers is quite natural. This extension is also reflected in the NCTM standards, since students are to learn the properties of addition and multiplication in elementary school, and then that these properties hold true for the real numbers is a natural extension.

There are several items we haven't checked yet from our axiom system for the real numbers. Namely, we need to check the following proposition:

Proposition 5.9 *If (A, B) is a Dedekind cut, then $(A, B) + (-(A, B)) = 0$ (by which we again mean for 0 to represent the equivalent cuts corresponding to 0).*

Proof: Let this sum be the cut (C, D) . Then $D = \{b + (-a) \mid b \in B, a \in A\}$. Since (A, B) a cut implies that $a < b$ for all $a \in A$ and $b \in B$, it follows that $b + (-a) > 0$ for all $a \in A$ and all $b \in B$. We need to show next, that if $x > 0$ and x is rational, then $x \in D$, as this would tell us that $D = \{x \mid x > 0\}$. Suppose $x > 0$ is a fixed rational number. Write this in lowest terms as $\frac{e}{f}$. Since (A, B) is a cut, there exists $b \in B$. Moreover, b is a rational number, and can thus be written as $\frac{g}{h}$ for some positive integers g and h . Since he and fg are positive integers, there exists a natural number n such that $n(he) > fg$. Consequently $nx > b$. Choose n to be the least positive integer such that $nx \in B$ (which exists by the Well Ordering Principle for the natural numbers). Then, $(n - 1)x \in A$ since $A \cup B = Q$. Moreover $nx + (-(n - 1)x) = x$ is an element of D . Thus D consists of all positive

rational numbers, and the cut $(C, D) = 0$ as desired.

Q.E.D. To establish this result, we again used fundamental properties of the integers.

To finish up our axioms, we need to show that the cut associated to 1 is a multiplicative identity, to define the multiplicative inverse of a non-zero cut, and to establish our completeness axiom that a non-empty set of cuts with a lower bound has a greatest lower bound.

For the first of these, note that if we look at the cut for 1 with 1 in the set B_1 , then establishing that (A_1, B_1) is a multiplicative identity is straightforward as given a cut (C, D) , then $(A_1, B_1) \cdot (C, D)$ is defined to be the cut (E, F) where

$$F = \{bd \mid b \in B_1, d \in D\}.$$

But $b \in B_1$ implies that $b \geq 1$. Thus $bd \geq d$ no matter what d is. Thus, if $d \in D$, by the definition of a cut, $bd \in D$. Therefore, $F \subseteq D$. On the other hand, $d = 1d \in F$ for all $d \in D$ since $1 \in B$. Thus $D \subseteq F$, and it follows that $D = F$, so that $(C, D) = (E, F)$ and we have the required result. (Note that we should have proven (A_0, B_0) really was an additive identity, but the proof of that is almost identical to the above proof.)

Thus, we need to define the multiplicative inverse. Again, we shall restrict to the case where $(A, B) > 0$. In this case, we define the multiplicative inverse of (A, B) by $(A, B)^{-1} = (C, D)$, where D is given by

$$D = \left\{ \frac{1}{a} \mid a \in A, a > 0 \right\},$$

and $C = Q \setminus D$. Note that we had to be a little careful in our definition of D as we could not simply choose for D the set of all inverses of elements of A , since that would have given us all negative numbers, and consequently, we would not have had a cut when we were done.

Proposition 5.10 *If $(A, B) > 0$ is a Dedekind cut, then $(A, B)^{-1}$ is a Dedekind cut, and $(A, B) \cdot (A, B)^{-1} \sim (A_1, B_1)$.*

Proof: We begin by showing the $(A, B)^{-1}$ is a cut. Let (C, D) be as above. First, as $(A, B) > 0$, we have that there exists a positive rational number $q \in A$. Thus $1/q \in D$, so that $D \neq \emptyset$. Since $0 \notin D$, it follows that $0 \in C$.

Hence both C and D are not empty. Note that if q is a rational number, then so is $1/q$. Consequently, since $A \subset Q$, it is also the case that $D \subset Q$. Now it is immediate that $Q = C \cup D$ and $C \cap D = \emptyset$ by the definition of C . Finally, suppose $d \in D$ and c is a rational number greater than d . Since $d \in D$, there exists a positive element $a \in A$ such that $d = \frac{1}{a}$. As $c > d > 0$, we have $0 < \frac{1}{c} < \frac{1}{d} = a$. Since (A, B) is a Dedekind cut and hence satisfies condition 3 for cuts, $\frac{1}{c} < a$ implies $\frac{1}{c} \in A$ (since it is rational). Moreover, as $\frac{1}{c} > 0$, it must be the case that $c = \frac{1}{\frac{1}{c}} \in D$. Thus, if $c > d$ is a rational number, then $c \in D$. By the contrapositive of this, we have that if $c \in C$ and $d \in D$ then $c < d$, establishing the third condition and showing that (C, D) is a cut.

It remains to show that $(A, B) \cdot (C, D) \sim (A_1, B_1)$. We will let (E, F) be the product of the cuts (A, B) and (C, D) . We will prove that $(E, F) \sim (A_1, B_1)$ by showing that if x is a rational number greater than 1 then $x \in F$ and if $x \in F$ then $x > 1$. Let $q \in F$. Then $q = bd$ for some $b \in B$ and some $d \in D$. As $d \in D$, we have $d = \frac{1}{a}$ for some positive $a \in A$. By condition 3 for the cut (A, B) , $a < b$ so that $b \cdot \frac{1}{a} > 1$. Thus $q \in F$ implies $q > 1$. Conversely, suppose $q > 1$ is a given rational number. Since $(A, B) > 0$, there exists $a \in A$ such that $a > 0$. Our goal is to use this a to find some $a' \in A$ and $b \in B$ such that $\frac{b}{a'} = q$, as this would force q to be an element of F . How do we find such an a' and b ? Noting that $\frac{b}{a'} < q$ is true if and only if $b = qa'$, it makes sense to look at the sequence

$$a, qa, q^2a, q^3a, \dots$$

At this point, we claim that for some $n \in N$, $aq^n \in A$ but $aq^{n+1} \in B$. To see this, note that by writing $q = 1 + \delta$, we have that $q^m > 1 + m\delta$ for any $m \in N$. Thus $aq^m > a + ma\delta$. Since $a\delta > 0$, by the Archimedean principle for the rational numbers (which follows from the well-ordering principle), we have for any $b \in Q$ there exists an $m \in N$ that $aq^m > b$. Choosing $b \in B$, we have for some $m \in N$ that $aq^m \in B$. Consequently, there is a least $n \in N$ such that $aq^n \in B$. Thus $aq^{n-1} \notin B$, so that $aq^{n-1} \in A$. Therefore, $q = aq^n \cdot \frac{1}{aq^{n-1}} \in F$. Since $q > 1$ was arbitrary, we have shown that F contains every rational number greater than 1, and consequently $(E, F) \sim (A_1, B_1)$ as desired.

Q.E.D.

The next of our axioms with a proof that requires more than definition chasing is the completeness axiom. We deal with this now. Recall that the

version of the completeness axiom we chose to deal with was the greatest lower bound law. (The other versions are the least upper bound principle and the requirement that Cauchy sequences converge.)

Theorem 5.11 *The set of Dedekind cuts satisfies the greatest lower bound principle.*

Proof: Let S be a non-empty set of Dedekind cuts that is bounded below. Since S is bounded below, there exists a cut (A, B) such that $(A, B) \leq (C, D)$ for all cuts $(C, D) \in S$. At this point, we are prepared to define our least upper bound. Let (U, V) be the pair of sets defined by

$$V = \cup_{(C,D) \in S} D,$$

and $U = Q \setminus V$. Our first goal is to show that (U, V) is a cut. Since S is nonempty, there exists some $(C, D) \in S$. As (C, D) is a cut, $D \neq \emptyset$, but $D \subset V$, so that $V \neq \emptyset$. On the other hand, let $a \in A$ be a rational number, and let $a' < a$ be another rational number. Since (A, B) is a cut, $a' \in A$. Let $v \in V$. By definition, this implies that $v \in D$ for some cut $(C, D) \in S$. If $(A, B) < (C, D)$ then $a' < v$. If $(A, B) \sim (C, D)$ on the other hand, as A and C differ by at most one element, that element must be greater than or equal to a . Consequently, $a' \in C$ again, and thus $a' < v$. Thus $a' < v$ for all $v \in V$. Thus $a' \in U$ and $U \neq \emptyset$. A straightforward argument shows that $U \cup V = Q$ and $U \cap V = \emptyset$. For the last condition, suppose $v \in V$ and $x > v$ be a rational number. By definition of V , it follows that $v \in D$ for some D where $(C, D) \in S$. As (C, D) is a cut, the last condition for cuts implies $s \in D$. Thus $x \in V$ also. Consequently, if $u \in U$ and $v \in V$, the condition that $U \cap V = \emptyset$ implies that $u < v$. Thus (U, V) is a cut.

We now claim that (U, V) is a greatest lower bound for S . First, note that $(C, D) \in S$ implies that $D \subseteq V$. Thus $(U, V) \leq (C, D)$, and (U, V) is a lower bound for S . Suppose that (E, F) is another lower bound for S . Let $x \in E$ and $v \in V$ be given. By definition $v \in D$ for some cut $(C, D) \in S$. As (E, F) is a lower bound for S , it follows that $x \leq v$. This, however, implies that $(E, F) \leq (U, V)$ as desired. Thus (U, V) is a greatest lower bound for S .

Q.E.D.

Thus, the Dedekind cuts are a model for the real numbers. The value of this particular model, is that it arises naturally from the number line, which

is the geometric motivation of the real numbers and many of our theorems about them. We note here that the more modern treatment of Dedekind cuts only considers one half of the cut. If you look back at our definitions of multiplication, addition, inverses, etc. you will notice that in every case we defined the lower half of the cut as the complement of the upper half of the cut. Consequently, we might have saved ourselves some trouble by simply defining cuts by their upper halves. This also allows for us to avoid the difficulty of having two equivalent cuts. The down side is that this one-sided definition is more complicated than the two sided definition, although this definition is hidden in our proofs for the most part.

At this point, we are finally ready to define the number represented by an arbitrary infinite decimal

$$\sum_{n=-k}^{\infty} a_n 10^{-n},$$

where $a_i \in \{0, 1, \dots, 9\}$. There are two ways to approach this right now. The standard technique, which works independent of whether we have used Dedekind cuts to define the real numbers or not, is to let

$$S = \left\{ \sum_{n=-k}^l a_n 10^{-n} \mid l = 1, 2, \dots \right\},$$

and to define the infinite decimal expansion as the least upper bound of the set S . Since we know that any non-empty set has a least upper bound by Theorem 5.1, every infinite decimal corresponds to a unique real number. That $1 = \overline{.9}$ then simply states that two slightly different sets have the same least upper bound, which seems reasonable since two ratios might also represent the same real number. On the other hand, we can explicitly use Dedekind cuts to define infinite decimals. That is, we define the real number associated to

$$\sum_{n=-k}^{\infty} a_n 10^{-n}$$

to be the Dedekind cut (A, B) , where

$$A = \left\{ x \in \mathbb{Q} \mid x \leq \sum_{n=-k}^l a_n 10^{-n} \text{ for some } l = 1, 2, \dots \right\},$$

and $B = \mathbb{Q} \setminus A$. Let us look at what happens in this case when we look at the cuts for 1 and for $\overline{.9}$. Writing (A, B) for the cut associated to $1 = 1.\overline{0}$,

we see that $A = \{x \leq 1 \mid x \in \mathbb{Q}\}$, while $B = \{x > 1 \mid x \in \mathbb{Q}\}$. Writing (C, D) for the cut associated to $.\overline{9}$, we see that $C = \{x < 1 \mid x \in \mathbb{Q}\}$, while $D = \{x \geq 1 \mid x \in \mathbb{Q}\}$. While these two cuts are not the same, they are **equivalent!** That is, the two decimals represent different but equivalent cuts. In this case, one can then see the different decimal representations of the number 1 as naturally corresponding to distinct but equivalent cuts. Thus, these two decimals correspond naturally to deciding which part of the line one wants to include the number 1 when creating the cut.

The preceding is a little disingenuous in that we might expect to see multiple decimal representations for all rational numbers, which we do not. However, we do have multiple representations for all $\frac{a}{b} \neq 0$ in base b (where a and b are integers with $b > 0$). Thus, the multiple representations has something to do with the base of the representation also.

Chapter 6

Classical Numbers

In this chapter, we shall discuss the important numbers e and π and the functions associated with them. While π is more commonly dealt with in the schools, the treatment here will begin with the number e . We have already briefly dealt with e on several occasions. In particular, in chapter 2, we gave several proofs that e is irrational. We also mentioned in chapter 4 that e is transcendental, and briefly discussed the complex plane. Similarly, we have already spent some time on π in the text, proving its irrationality and stating that it too is transcendental. Our goal in this chapter is to discuss the historical development of the logarithmic function and the function e^x . After discussing e , our next goal will be to discuss the number π , spending a little time on how one finds the billions of digits of π that we know today. Afterwards, we shall spend some time discussing the trigonometric functions and how one might define them in a more natural way in the high school classroom. As always, Klein's book *Elementary Mathematics from an Advanced Standpoint* is a major basis for our discussion.

6.1 The Logarithmic Function

In the schools, the logarithmic functions are naturally introduced as the inverse function of the exponential function $y = b^x$ where b is a fixed positive number, and we write $x = \log_b(y)$. To do so, we begin by restricting x to the positive integers, and then slowly building up values for b^x when x is a negative number, and then a rational number. At each step, we do so by appealing to the multiplicative formula $b^{a+c} = b^a b^c$. As we get to rational

values for x , we recognize that for all rational values to make sense, we must restrict b to a positive number. At this point, we extend the function $y = b^x$ to irrational numbers as necessary using the density of the rational numbers. Of course, these texts never mention the idea that one could have allowed b to be a negative number and simply restricted the rational numbers to quotients of two odd integers, at which point, we could have appealed to the density of these numbers to define the logarithm for all numbers. They avoid this because to do so calls into play complex analysis, a topic well beyond the secondary student. Rather, we simply inform students that b must be positive for the logarithm to make sense.

The natural logarithmic function is often introduced as the inverse function of the exponential function e^x where e is the *natural base* of the logarithmic function, the number 2.718281828459045... To do this, the standard text uses the idea of compound interest. The idea being that if one invests P dollars in the bank and it compounds monthly at 100% interest (12 times a year), then at the end of one year, you have $P(1 + \frac{1}{12})^{12}$ dollars. Alternatively, if you compound the interest daily (banks use 360 days a year) you have $P(1 + \frac{1}{360})^{360}$ dollars. Continuing in this fashion, if you compound it *instantaneously*, then at the end of one year you have

$$P \cdot \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$$

dollars. At this point, we define the number e to be the limit. One can then define e^x similarly as

$$e^x = \lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n.$$

Unfortunately, this definition does not do much for students attempting to gain an understanding of the number e and why it should arise mathematically. Let us turn to the historical development of the number e and the logarithmic function so that we can gain a better understanding.

The first logarithmic tables were published by the Scottish mathematician John Napier (1550-1617) in 1614. His methods for calculating these tables appeared after his death in 1619. For the basis of his logarithmic function, Napier chose the number .999999999, as he was interested in applying logarithms to trigonometric functions, where one deals primarily with numbers less than 1 (and Napier's choice made the logarithms of such numbers positive). The Swiss mathematician Jobst Bürgi published his set of log tables in 1620 independently of Napier using the base $b = 1.0001$.

Both Napier and Bürgi calculated their logarithms by trying to solve the equation $x = b^y$ for integer values of y , and try to find an arrangement where the known values of these numbers are as close together as possible. In the modern treatment, we allow for y to be a rational number, thus squeezing together the values of b^y . Napier and Bürgi, however, both hit upon the useful idea of taking b very close to 1, so that for integer values of y , the numbers x are close together. Even so, calculations would appear to be very complicated. (Try raising 1.0001 to the 5000th power by hand.) However, using the idea of difference methods, if we take Bürgi's base, we can work inductively. That is, if we know that $x = (1 + 10^{-4})^y$, then letting

$$x + \Delta x = (1 + 10^{-4})^{y+1} = x(1 + 10^{-4}),$$

we obtain that $\Delta x = \text{frac}x10^4$. Letting $\Delta y = 1$ (the difference in the y values in this case, we obtain the difference equation (or essentially an approximation of the derivative)

$$\frac{\Delta y}{\Delta x} = \frac{10^4}{x}.$$

Thus, if we know the $\log_b(x)$, then

$$\log_b(x + \delta) = \log_b(x) + \delta \frac{10^4}{x}.$$

In fact, this is essentially how Bürgi did his calculations. In a similar way, Napier's logarithms satisfy a difference equation of

$$\frac{\Delta y}{\Delta x} = -\frac{10^7}{x}.$$

At this point it is worthwhile to notice that if we take as our base $(1.0001)^{10000}$, we only change the decimal in the logarithm. Moreover, the new base in this case is 2.718146, which is extremely close to e !

At this point, we should take a geometric look at what we have. In particular, taking the function $y = \frac{1}{x}$, and starting at $x = 1$, we arrive at a sum, $z = \sum \frac{\Delta x}{x}$, where z is the value of the logarithm, and x and Δx change throughout the sum so that the rectangle having height $\frac{1}{x}$ (the y value on the curve at x) and width Δx . At this point the picture should remind us of the approximations of the integral, and in fact, one can define the natural logarithm by

$$\ln(a) = \int_1^a \frac{1}{x} dx.$$

This is the historical treatment and the main step was taken in 1650, when the infinitesimal calculus was attacking the problems of the area under (or the *quadrature* of) various curves. Nicolas Mercator (1620-1687) was in the forefront of making this definition and is responsible for the name “natural logarithm.” *Revisiting the History of the Logarithm* by John Fauvel [2]. Indeed, much of this material is taken from there.)

Before examining this definition more carefully, we will take a brief detour into the history of calculations with large numbers. The English mathematician Henry Briggs (1566-1630) recognized the calculational value of having logarithms base 10, and it was he who first introduced them in 1617. (For a beautiful treatment of this material, I highly recommend the article In fact, an interesting sidelight here is that before logarithms were invented, mathematicians used trigonometry to calculate large products ([10], [11]) with the method of prosthaphaeresis. This method converts the problem into one of addition, using sines and cosines. The basic idea is to use the identities

$$\begin{aligned} 2 \cos A \cos B &= \cos(A - B) + \cos(A + B) \\ 2 \sin A \sin B &= \cos(A - B) - \cos(A + B) \\ 2 \sin A \cos B &= \sin(A - B) + \sin(A + B) \\ 2 \cos A \sin B &= \sin(A + B) - \sin(A - B). \end{aligned}$$

To see better how this might work, let us take an example. Suppose you want to multiply 1023 by 3101 using this method. The first step is to think of a large circle with radius 10000. To use the first of the above equations, we need to express 1023 and 3101 using our radius and cosines and sines. For this case, we will look at the fractions

$$\begin{aligned} \frac{1023}{10000} &= .1023 = \cos(84.128\dots) \\ \frac{3101}{10000} &= .3101 = \cos(71.9347\dots). \end{aligned}$$

The first equation then tells us that

$$\begin{aligned} 2 \cdot \frac{1023}{10000} \cdot \frac{3101}{10000} &= \cos(84.128\dots - 71.9347\dots) + \cos(84.128\dots + 71.9347\dots) \\ &= \cos(12.19\dots) + \cos(156.06\dots) \\ &= .0634818722\dots \end{aligned}$$

The above answer is approximate due to roundoff error, however, we now get from here that $1023 \cdot 3101 = \frac{1}{2} \cdot 100000000 \cdot 0634818722 \dots = 3174093.6129 \dots$. Since the correct result is 3172323, we see that the round off error from the calculations has had some effect, but we are still correct to within 4%. While this isn't too bad, one ought to think about whether we could have done better without making our job too much harder. Mathematicians had extensive trigonometric tables (which were created for the purpose of calculating), and these often went further than simply four digits. Thus, they could ask for more digits of accuracy. Doing our calculations in radians and asking for 11 decimal places, we have:

$$\begin{aligned} 2 \cdot \frac{1023}{10000} \cdot \frac{3101}{10000} &= \cos(1.46831704802 - 1.25549811094) + \cos(1.46831704802 + 1.25549811094) \\ &= \cos(.21281893708) + \cos(2.72381515896) \\ &= .97743939425 + (-.91399293425) \cdot 06344646. \end{aligned}$$

Multiplying this answer by 50000000 yields the correct answer 3172323, in fact, we should only need to require 8 decimal place accuracy as the integer part of the answer will correspond to the first 9 decimals. The trick in the mix is that we need to be sure that if the angle is accurate to 9 decimal places then the cosine of the angle is accurate to 8 decimal places. For this, we can turn to calculus and it follows from a derivative check (see homework).

Turning back to the definition of e , there can be other ways to get at this funny number. The connected math series, for example, suggests looking at slopes of exponential functions at $x = 0$. In this case, we are looking at finding a number a such that the slope of the curve a^x is 1 for $x = 0$. This corresponds to the limit question: for what a does $\lim_{x \rightarrow 0} \frac{a^x - 1}{x} = 1$. One can graphically evaluate this limit for various choices of a and fairly quickly see that the correct value is somewhere between 2.6 and 2.8. Put in small values for x and graph the result on a . Using the intersect command with the function $y = 1$ and $y = \frac{a^{.00001} - 1}{.00001}$, you can get an even better method of approximating e nailing down many digits.

A fourth way of defining e comes from using Taylor's series. Since the exponential function is defined as the function $f(x)$ that has the property that $f'(x) = f(x)$, and $f(0) = 1$, (which can be shown to be the same as stating that e is the number such that e^x has slope 1 at $x = 0$ (see below), then it follows that if $f(x)$ has a Taylor's series that converges at a neighborhood of

$x = 0$, then it must be of the form

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Consequently, $f(1) = \sum_{n=0}^{\infty} \frac{1}{n!}$. This series quickly converges to e (after 10 terms, the first 6 decimal places are correct). All that is left is to show that $f(x) = a^x$ and then we would be set.

Theorem 6.1 *Let $f(x)$ be a function such that $f(0) = 1$ and $f'(x) = f(x)$ for all x , then $f(x) = e^x$, where e is the real number such that e^x has derivative 1 at 0.*

Proof: We will begin by showing that if e has the property that the derivative of e^x is 1 at $x = 0$, then $f(x) = e^x$. To do this, we begin by showing that the derivative of $g(x) = e^x$ is e^x at all x . Calculating the derivative we have

$$\begin{aligned} F'(x) &= \lim_{\delta \rightarrow 0} \frac{e^{x+\delta} - e^x}{\delta} \\ &= \lim_{\delta \rightarrow 0} \frac{e^x e^\delta - e^x}{\delta} \\ &= e^x \lim_{\delta \rightarrow 0} \frac{e^\delta - 1}{\delta} \\ &= e^x. \end{aligned}$$

The next step is to show that there is only one function (i.e., the function is unique) such that $g'(x) = g(x)$ and $g(0) = 1$. In this case, we simply look at the difference of the two functions. At this point, we will quote a couple of theorems we won't prove, namely that Taylor's series are infinitely differentiable functions within their radius of convergence (i.e., at all values x where the sum makes sense on an open interval containing x) and that exponential functions are also infinitely differentiable. (At some level, this is really cheating since the property of being infinitely differentiable (or C^∞) is equivalent to equaling the Taylor's series in a given interval, but let's ignore this for now.) Given this, the function $F(x) - f(x)$ is necessarily infinitely differentiable. Moreover, $F(0) - f(0) = 0$ and $(F(x) - f(x))' = F(x) - f(x)$. Since the result would follow if we show that $F(x) - f(x) = 0$ for all x , we have changed our original goal to showing that if $g(x)$ is infinitely differentiable,

$g(0) = 0$ and $g'(x) = g(x)$, then $g(x) = 0$ for all x . Suppose $g(a) \neq 0$ for some positive a (the case for negative a works similarly). If $g(a - x) > 0$ for all $x \in (0, .5)$, then we replace a by $a - .25$. Continuing in this manner, we may assume that $g(a - x) = 0$ for some $x \in (0, .5)$. Let

$$S = \{z \in (0, a) \mid g(z) = 0\},$$

and let c be the least upper bound of S . By the intermediate value theorem, there exists a $y_0 \in (c, a)$ such that $g'(y_0) = \frac{g(a) - g(c)}{a - c}$. That is $g(y_0) = g'(y_0) = \frac{g(a)}{a - c} > 2g(a)$ as $0 < a - c < .5$. Repeating the same argument with y_0 in place of a , there exists $y_1 \in (c, y_0)$ such that $g(y_1) = g'(y_1) = \frac{g(y_0)}{y_0 - c} > 4g(a)$. Continuing in this manner, we have $y_k \in (c, a)$ such that $g(y_k) > 2^k g(a)$. This, implies that $g(x)$ is unbounded on the interval (c, a) , contradicting the continuity of $g(x)$. Hence $g(a)$ is not positive. A similar argument eliminates the case that $g(a)$ is negative (do this!). Therefore, $g(x) = 0$ for all x , and hence $f(x) = F(x)$ as desired.

Q.E.D.

Bibliography

- [1] Chrystal, G., *Algebra*, Adam and Charles Black, London, 2nd edition, 1900.
- [2] Faubel, J., Revisiting the History of Logarithms, *Learn from the Masters* Swetz, Fauvel (eds.), Mathematics Association of America, 1995, pp. 39-48.
- [3] Gelfond, A.O., *Doklady Akad. Nauk S.S.S.R.*, **2**, 1934, pp. 1-6.
- [4] Hille, E., *American Mathematical Monthly*, **49**, 1942, pp. 654-661.
- [5] Klein, F., *Elementary Mathematics from an Advanced Viewpoint*, Dover, New York (translated into English by E.R. Hedrick and C.A. Noble).
- [6] Klein, F., *et al.*, *Famous Problems and Other Monographs*, Chelsea, New York, 1962.
- [7] Niven, I., A simple proof that π is irrational, *American Mathematical Monthly*, 1947, p. 509.
- [8] Niven, I., *Irrational Numbers*, The Carus Mathematical Monographs, No. 11, John Wiley and Sons, New Jersey, 1956.
- [9] Polya, G., *How To Solve It*, second edition, Doubleday, New York, 1957.
- [10] Pierce, R.C., Sixteenth-century astronomers had prosthaphaeresis, *Mathematics Teacher*, **70**, 1977, pp.613-614.
- [11] Resnikoff, H., and Wells, R., *Mathematics in Civilization*, Holt, Rinehart, and Winston, New York, 1973.

- [12] Schneider, T., *J. Reine Angew. Math.*, **172**, 1935, pp. 65-69.
- [13] Schoenfeld, A., *Mathematical Problem Solving*, Academic Press, Orlando, 1985.
- [14] Sfard, A., On reform movements and the limits of mathematical discourse, *Mathematical Thinking and Learning*, Vol. 2 **3**, 2000, pp. 157-190.
- [15] Smith, D., Latham, M., *The Geometry of René Descartes: translated from the French and Latin*, Open Court Publishing Company, Chicago, 1925.
- [16] Stewart, I., *Galois Theory*, second edition, Chapman & Hall, 1989.
- [17] Swertz, F., Fauve., j., et al., *Learn from the Masters*, Mathematical Association of America, 1995.
- [18] Uhl, J.