Survey of Modern Mathematical Topics inspired by History of Mathematics

Paul L. Bailey

Department of Mathematics, Southern Arkansas University
E-mail address: plbailey@saumag.edu
Date: January 21, 2009

Contents

Preface

Divide each chapter into three parts:
1) Time and Place
2) Personalities
3) Mathematical Topic(s)

Chapter - Ancient Mathematics
Time and Place: Prehistory, Egyptian, Babylonian
Personalities: none
Topics: Bases

Chapter - Early Greek

Chapter - Analytic Geometry
Time and Place: France early 17th century
Personalities: Descartes, Fermat, Pascal
Topic: Analytic Geometry, Tangents

CHAPTER I

Bases

1. Introduction

Consider the difference between the words *number* and *numeral*, as they are used by mathematicians.

Webster's New World dictionary defines number as *a symbol or word, or a group of either of these, showing how many or which one in a series.* This is clearly not what we mean when we refer to rational or real numbers. Yet, the alternate definitions are even further from our usage. Perhaps closer would be *an idea corresponding to a quantity.* Let's take that for now (although it certainly seems to exclude complex numbers).

Webster's does a better job with the second word, defining numeral as *a figure, letter, or word, or a group of any of these, expressing a number.* So if a number is an idea, a numeral is an expression of an idea.

Our standard way of writing numbers depends on the choice of 10 as a base; this is called the *decimal system.* For example, the number eight thousand six hundred forty two divided by twenty five is written in decimal as

$$\frac{8642}{25} = 345.68 = 3(10^2) + 4(10^1) + 5(10^0) + 6(10^{-1}) + 8(10^{-2}).$$

However, the choice of ten is arbitrary, and other cultures have made other choices.

In this note, we explore how to express numbers in differing bases, and discover an interesting fact about radix expansions in alternate bases.

2. Integer Expansion Algorithm

The property of the integers which is pivotal is understanding bases is the way an integer breaks down into a quotient and remainder when it is divided by another integer. We state the result we use.

Proposition 1. Division Algorithm
Let $m, n \in \mathbb{Z}$. There exist unique integers $q, r \in \mathbb{Z}$ such that

$$n = mq + r \qquad \text{and} \qquad 0 \leq r < m.$$

We call q the *quotient* and r the *remainder*.

Recall that a polynomial is a function of the form

$$f(x) = \sum_{i=0}^{k} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k,$$

where the coefficients a_i are selected from some prespecified set. We will use the division algorithm to show how to express an integer n as a polynomial in b, where b is the base. That is, for $b, n \in \mathbb{Z}$ with with $b \geq 2$, we find f as above with and $0 \leq a_i < b$ such that $f(b) = n$.

Lets first consider how we compute $f(b)$. The naive way to evaluate the polynomial f at a given value for x involves evaluating each monomial separately and adding the values together. This requires k additions and $\sum_{i=1}^{k} i = \frac{k(k+1)}{k}$ multiplications.

However, we may factor the polynomial thusly:

$$f(x) = a_0 + x(a_1 + x(a_2 + \cdots + x(a_{k-1} + x(a_k)) \dots)).$$

Evaluating this at the same x requires k additions and k multiplications.

Proposition 2. Integer Expansion Algorithm
Let $b, n \in \mathbb{Z}$ with $b \geq 2$. Then there exists a unique polynomial

$$f(x) = \sum_{i=0}^{k} a_i x^i$$

with integer coefficients such that

(1) $f(b) = n$;
(2) $0 \leq a_i < b$, *with* $a_k > 0$.

We call the coefficients a_i the *base b digits* of the number n. We may compute these as follows. Let $n \in \mathbb{Z}$; for simplicity assume n is positive. The division algorithm states that $n = bq + r$ for some $q, r \in \mathbb{Z}$ with $0 \leq r < b$. That $0 \leq r < b$ states that r is a digit in base b.

Set $q_0 = n$, $q_1 = q$, and $r_0 = r$ so that the above equation becomes

$$q_0 = bq_1 + r_0.$$

Then inductively compute

$$q_i = bq_{i+1} + r_i.$$

Since the q_i's are positive and decreasing, this process eventually ends, say at the k^{th} stage, so that

$$q_k = bq_{k+1} + r_k \quad \text{with} \quad q_{k+1} = 0;$$

at this point, $r_k = q_k$. If we plug this back into the previous equation $q_{k-1} = bq_k + r_{k-1}$, we see that $q_{k-1} = br_k + r_{k-1}$, which we rewrite as $q_{k-1} = r_{k-1} + br_k$. If we then take this and plug it back into its predecessor and rearrange, we obtain $q_{k-2} = bq_{k-1} + r_{k-2} = r_{k-2} + b(r_{k-1} + br_k)$. Next, and in the same manner, we find that $q_{k-3} = r_{k-3} + b(r_{k-2} + b(r_{k-1} + br_k))$. Continuing this process, we eventually arrive at

$$n = q_0 = r_0 + b(r_1 + b(r_2 + b \dots (r_{k-1} + br_k) \dots)).$$

Rewritten in standard polynomial form, using summation notation, this becomes

$$n = \sum_{i=0}^{k} r_i b^i.$$

In shortened notation, the base b numeral representing the number n is written

$$n = (r_k r_{k-1} \dots r_1 r_0)_b.$$

That is, the digits of n written in base b are the remainders upon successive division by b.

3. Radix Expansion Algorithm

The expression of a real number in base b is called its *base b radix expansion*. We show how to find this for a real number between 0 and 1; combine this with the integer expansion algorithm to find the base b expansion of any real number.

Definition 3. A *power series* is a function of the form

$$f(x) = \sum_{i=0}^{\infty} a_i x^i,$$

where $a_i \in \mathbb{C}$.

For example, $|x| < 1$ and $a_i = 1$ for all i, then the power series is a convergent geometric series.

Proposition 4. Radix Expansion Algorithm
Let $z \in \mathbb{R}$ with $0 < z < 1$ and $b \in \mathbb{Z}$ with $b \geq 2$. Then there exists a unique power series

$$f(x) = \sum_{i=0}^{\infty} a_i x^i$$

with integer coefficients such that

(1) $f(\frac{1}{b}) = z$;
(2) $0 \leq a_i < b$ *for all* i;
(3) *if* $a_i = b - 1$ *then there exists* $j > i$ *such that* $a_i \neq b - 1$.

Note that since $a_i \leq b - 1$ for all i, then $f(\frac{1}{b}) \leq \sum_{i=0}^{\infty} (\frac{b-1}{b})^i$, which is a geometric series and therefore is convergent. Thus $f(\frac{1}{b})$ also converges.

Let $z \in (0, 1)$. Then $0 \leq bz_0 < b$. Multiply z_0 by b and take the integer part; call this p_1. Set

$$z_1 = bz_0 - p_1 \text{ with } p_1 \in \mathbb{Z}, 0 \leq p_1 < b, \text{ and } 0 \leq z_1 < 1.$$

Repeat this: $z_2 = bz_1 - p_2$, $z_3 = bz_2 - p_3$, and so forth. Inductively, take z_i and produce p_{i+1} and z_{i+1} such that

$$z_{i+1} = bz_i - p_{i+1} \text{ with } p_{i+1} \in \mathbb{Z}, 0 \leq p_{i+1} < b, \text{ and } 0 \leq z_{i+1} < 1.$$

The base b radix expansion of z is the series

$$z = \sum_{i=0}^{\infty} p_i \frac{1}{b^i}.$$

For the valiant reader, we explain why the series above converges to z. To do this, we show that the difference between the z and the partial sums of the series becomes as small as we want as we add additional terms. Such proofs often begin with the phrase "let $\epsilon > 0$"; this means that ϵ is arbitrarily small, and we show that the difference eventually becomes less than ϵ.

Let $\epsilon > 0$ and select $k \in \mathbb{N}$ so large that $\frac{1}{b^k} < \epsilon$. Then $\frac{z_{k+1}}{b^{k+1}} < \epsilon$. Solve each equation $z_{i+1} = bz_i - p_{i+1}$ for z_i to obtain

$$z_i = b^{-1}(p_{i+1} + z_{i+1}).$$

Rewind all this by substituting each such equation into the previous one:

$$\begin{aligned} z_k &= b^{-1}p_{k+1} + b^{-1}z_{k+1}; \\ z_{k-1} &= b^{-1}(p_k + b^{-1}p_{k+1}) + b^{-2}z_{k+1}; \\ z_{k-2} &= b^{-1}(p_{k-1} + b^{-1}(p_k + b^{-1}p_{k+1})) + b^{-3}z_{k+1}; \end{aligned}$$

and so forth, until eventually

$$z = z_0 = b^{-1}(p_1 + b^{-1}(p_2 + b^{-1}(\dots b^{-1}(p_k + b^{-1}p_{k+1})\dots))) + b^{-(k+1)}z_{k+1}.$$

Thus

$$z - \sum_{i=0}^{k} p_i \frac{1}{b^i} = \frac{z_{k+1}}{b^{k+1}} < \epsilon,$$

which shows the convergence we desire.

4. Rational Expansion Property

Let $z \in \mathbb{Q}$, and for simplicity assume that $0 < z < 1$. Then $z = \frac{m}{n}$ for some $m, n \in \mathbb{N}$ with $m < n$ such that $\gcd(m, n) = 1$; this last condition guarantees that n is as small as possible.

We may obtain the base b radix expansion for z,

$$z = \sum_{i=0}^{\infty} p_i \frac{1}{b^i},$$

by repeated use of the division algorithm; this is the normal process of division, in base b, dividing n into m. Since $m < n$, we must first multiply m by b; then the quotient will be p_1 and the remainder will be an integer which is less than n:

$$bm = np_1 + r_1.$$

Next we multiply r_1 by b and divide, to get

$$br_1 = np_2 + r_2.$$

Inductively find p_i and r_i such that

$$br_i = np_{i+1} + r_{i+1}.$$

Now at each stage, $r_i < n$, so eventually two of remainders will be the same; let k be the smallest integer such that

$$r_k = r_i$$

for some $i < k$. Then $p_{i+j} = p_{k+j}$ for $j = 1, \ldots, k-i$, and this pattern continues to repeat. We call this a radix expansion whose repeating part starts after the i^{th} place and has length $k - i$.

On the other hand, if $z = \sum_{i=0}^{\infty} p_i \frac{1}{b^i}$ is a radix expansion whose repeating part starts after the i^{th} place of length $k - i$, then $(b^k - b^i)z$ is an integer, and

$$z = \frac{(b^k - b^i)z}{b^k - b^i}$$

expresses z as a rational number.

Together, we see that

Proposition 5. Rational Expansion Property *Let $z \in \mathbb{R}$, with $0 < z < 1$. Then the base b radix expansion of z repeats if and only if $b \in \mathbb{Q}$. Moreover, if $z = \frac{m}{n}$, then the sum of the lengths of the nonrepeating and the repeating parts of the radix expansion of z is less than or equal to n.*

If the repeating part of the base b radix expansion of z consists of a single repeating zero, we say that it *terminates*.

5. Regular Numbers

Definition 6. Let $n \in \mathbb{Z}$ with $n \geq 2$. We say that n is *base b regular* if the base b radix expansion of its reciprocal terminates.

Proposition 7. *Let $n \in \mathbb{Z}$ with $n \geq 2$. Then n is base b regular if and only if n is a product of powers of the prime divisors of b.*

Proof. We prove both directions of the implication.

($\Rightarrow$) Suppose that n is base b regular and that p is a prime divisor of n. We show that b is a prime of divisor of b, so that all primes in n are in b, and n must be the product of prime divisors of b.

Since n is base b regular, $\frac{1}{n}$ has a finite base b radix expansion, say of length i. Then $\frac{b^i}{n}$ is an integer, and n divides b^i. That is, b^i is a multiple of n, so every prime divisor of n must also be a prime divisor of b^i, and therefore of b itself.

($\Leftarrow$) Suppose that n is a product of powers of the prime divisors of b. Then for some $k \in \mathbb{N}$, we have $n \mid b^k$, say $nm = b^k$. Then

$$\frac{1}{n} = \frac{m}{b^k},$$

which clearly has a finite base b radix expansion. □

6. Problems

Problem 1. Convert from base 10 to the given base.

(a) Write 1234 in base 5.
(b) Write 1234 in base 7.
(c) Write 1234 in base 20.

Problem 2. Convert to base 10.

(a) $(1234)_5$
(b) $(1234)_7$
(c) $(1234)_{20}$

Problem 3. Solve the following equations for the positive integers n and b.

(a) $n = (13425)_b = (4115)_{2b}$
(b) $n = (1234)_b = (532)_{2b-1}$

CHAPTER II

Constructibility

ABSTRACT. We discuss the classical Greek notion of constructibility of geometric objects. The reader is invited to obtain a ruler and compass to perform the exercises and follow the constructions described in the proofs.

1. Construction with Straight-Edge and Compass

The drawings of the ancient Greek geometers were made using two instruments: a straight-edge and a compass.

A *straight-edge* draws lines. With the straightedge, we are permitted to draw a straight line of indefinite length through any two given distinct points. The straight-edge is unmarked; it cannot measure distances.

A *compass* draws circles. With the compass, we are permitted to draw a circle with any given point as the center and passing through any given second point. The compass collapses if it is lifted; we are not *a priori* permitted to use it to measure the distance between given points, and draw a circle around another given point of the same radius.

The straight-edge and the compass have come to be known as *Euclidean tools*, although the quest to construct points using them pre-dates Euclid by two centuries.

2. Construction of Points in a Plane

Let P denote the set of all points in a plane, and let $Q \subset P$.

A line in P is *given by* Q if there exist two points in Q which lie on P.

A circle in P is *given by* Q if the center of the circle is in Q, and there exists a point in Q which lies on the circle.

A point $A \in P$ is *immediately constructible* from Q if one of the following hold:

(a) $A \in Q$;
(b) A is the point of intersection of two lines which are given by Q;
(c) A is a point of intersection of a line and a circle which are given by Q;
(d) A is a point of intersection of two circles which are given by Q.

A point $A \in P$ is *eventually constructible* from Q if there exist a finite sequence of points $A_1, A_2, \ldots, A_n$ such that $A = A_n$ and for $j = 1, \ldots, n$, A_{j+1} is immediately constructible from $Q \cup \{A_1, \ldots, A_j\}$.

3. Standard Constructions

Let P denote a plane. For $A, B \in P$, define the following:

- AB is the line in P through A and B;
- $\overline{AB}$ is the line segment between A and B;
- $|AB|$ is the distance between A and B;
- $A - B$ is the circle through B with center A.

Also, if $C, D \in P$, then $AB \parallel CD$ represents the statement that line AB is parallel to line CD, and $AB \perp CD$ represents the statement that line AB is perpendicular to line CD.

Let Q be a set of points in the plane. We say that a line segment is constructible from Q if its endpoints are constructible from Q

Proposition 1. *Given points A and B, it is possible to construct the midpoint Z of $\overline{AB}$.*

Construction. We are given A and B.

(a) Let C and D be the points of intersection of circle $A - B$ and circle $B - A$.

(b) Let Z be the intersection of line AB and line CD.

Then Z is the midpoint of $\overline{AB}$. □

Proposition 2. *Given points A and B, it is possible to construct a point Z such that $AB \perp BZ$.*

Construction. We are given A and B.

(a) Let C be the point of intersection of line AB and circle $B - A$ which is not A.

(b) Let Z be one of the points of intersection of circle $A - C$ and circle $C - A$.

Then $AB \perp BZ$. □

Proposition 3. *Given noncolinear points A, B, and C, it is possible to construct a point Z on the line AB such that $AB \perp CZ$.*

Construction. We are given A, B, and C. If $CB \perp AB$, let $Z = C$. Otherwise, construct Z as follows.

(a) Let D be the point of intersection of line AB and circle $C - B$ which is not B.

(b) Let Z be the midpoint of $\overline{BD}$.

Then $AB \perp CZ$. □

Proposition 4. *Given noncolinear points A, B, and C, it is possible to construct a point Z such that $AB \parallel CZ$.*

Construction. We are given A, B, and C.

(a) Let D be the point of intersection of line AB and the line through C which is perpendicular to line AB.

(a) Let Z be the point of intersection of the line through A which is perpendicular to line AB and the line through C which is perpendicular to line CD.

Then $AB \parallel CZ$. □

4. Transference of Distance

Suppose we are given points A, B, and C. A *modern compass* is capable of holding its shape when lifted from the page, so that the distance between A and B can be measured using the modern compass, and then the compass is set down on C to draw a circle with center C and radius $|AB|$. We may call this process *transference of distance*. The Euclidean compass is not *a priori* capable of this feat; however, we can prove that this construction is possible. We do this by constructing a parallelogram $ABCZ$, so that $|AB| = |CZ|$.

Proposition 5. *Given noncolinear points A, B, and C, it is possible to construct a point Z such that polygon $ABCZ$ is a parallelogram.*

Construction. We have points A, B, and C.

(a) Let Z be the point of intersection of the line through C parallel to AB, and the line through A parallel to BC.

□

5. The Three Greek Problems

As the Greeks investigated what could be accomplished with their Euclidean tools, three interesting unsolved problems arose.

Greek Problem 1 (Duplication of the Cube)**.** Given a cube, construct a cube with double the volume.

Greek Problem 2 (Trisection of an Angle)**.** Given an angle, construct an angle one third as large.

Greek Problem 3 (Quadrature of the Circle)**.** Given a circle, construct a square with the same area.

We now attempt to make the statements of these problems precise, using modern notation.

6. Construction of Squares

A square is constructible if its vertices are constructible.

Quadrature is the process of constructing a square whose area is equal to the area of a given plane region. A plane region with area x is called *quadrable* if it is possible to construct a square with area x. By the Proposition 2, this is equivalent to the the ability to construct a line segment of length $\sqrt{x}$.

The ancient Egyptians estimated areas of certain regions; for example they estimated that the square on 8/9 of the diameter of a circle is its quadrature. The area x of the circle with radius r would then be approximately

$$x \approx \left(\frac{8}{9}(2r)\right)^2 = \frac{256}{81}r^2;$$

this produces $\pi \approx 3.16049$.

The ancient Greeks concentrated on discovering which regions were precisely quadrable, via construction with Euclidean tools.

The third Greek problem asks if a given circle is quadrable.

7. Construction of Angles

Let P denote a plane. For $A, B, C \in P$, define the following:

- $\angle ABC$ is the angle between the line segments $\overline{AB}$ and $\overline{BC}$.

We say that an angle α is constructible from $Q \subset P$ if it is possible to construct points A, B, and C from Q such that $\alpha = \angle ABC$.

To say that an angle α is given; means that we are given points A, B, and C such that $\alpha = \angle ABC$. A *bisector* of this angle is a line BD such that $\angle ABD = \angle DBC$; then necessarily $\angle ABD = \frac{\alpha}{2}$.

Proposition 6. *Given an angle $\angle ABC$, it is possible to construct a point Z such that $\angle ABZ = \angle ZBC = \frac{\angle ABC}{2}$.*

Construction. We are given A, B, and C, with B as the vertex of the angle.

(a) Let D be the point of intersection of BC and $B - C$.

(a) Let Z be the midpoint of $\overline{CD}$.

Then $\angle ABZ = \angle ZBC$. □

Thus every given angle is *bisectable*; the second Greek problem asks if every given angle is *trisectable*.

8. Construction of Points in Space

Let S denote the set of all points in three dimensional space, and let $A, B \in S$. Although the line through A and B is well defined, there are many circles in space whose center is A which pass through B. We do not wish to say that all such circles are constructible.

We say that a plane $P \subset S$ is constructible from a set $Q \subset S$ if there exist three noncolinear points in Q which lie on P. Now circles are constructible from Q if we may construct the plane on which they lie. This gives meaning to the notion of constructibility of a point in space.

A cube is constructible if it is possible to construct its vertices in space.

The first Greek problem asks if, given a cube in space, it is possible to construct a cube in space whose volume is double that of the given cube. This is equivalent to asking if, given a line segment whose length is that of a side of the original cube, it is possible to construct a line segment whose length is that of a cube with double the volume.

9. Construction of Real Numbers

Let P be a plane and let $Q \subset P$. Let $x \in \mathbb{R}$. We say that x is constructible from Q if a line segment whose length is $|x|$ is constructible from Q. Moreover, we say simply that x is a *constructible real number* if x is constructible from $\{A, B\}$ for some $A, B \in P$ with $|AB| = 1$. Since we may consider a point to be a line segment of length 0, we consider 0 to be a constructible number.

Proposition 7. *Let $x, y \in \mathbb{R}$ be constructible. Then $x + y$ is constructible.*

Construction. Since x and y are constructible, it is possible to construct line segments of length $|x|$ and $|y|$. By Proposition 5, it is possible to construct a circle of radius $|y|$ centered at any given point.

(a) Let A and B be points such that $|AB| = |x|$.

Case 1 First assume that x and y have the same sign.

(b) Let Z be the point of intersection of line AB and the circle centered at B of radius y such that B lies on $\overline{AZ}$.

Now $\overline{AZ}$ is a line segment of length $|x| + |y| = |x + y|$.

Case 2 Next assume that x and y have different signs, and without loss of generality assume that $|x| \geq |y|$.

(b) Let Z be the point of intersection of line AB and the circle centered at B of radius y such that Z lies on $\overline{AB}$.

Now $\overline{AZ}$ is a line segment of length $|x| - |y| = |x + y|$. □

Proposition 8. *Let $x \in \mathbb{R}$ be constructible. Then $-x$ is constructible.*

Reason. This follows immediately from the definition. □

Proposition 9. *Let $x, y \in \mathbb{R}$ be constructible. Then xy is constructible.*

Construction. Since 1, x and y are constructible, it is possible to construct line segments of length 1, $|x|$, and $|y|$. Without loss of generality, we may assume that x and y are positive.

- **(a)** Let A and B be points such that $|AB| = x$.
- **(b)** Let C be a point of intersection of the line through A which is perpendicular to line AB and a circle centered at A of radius 1.
- **(c)** Let D be the point of intersection line through AC and the circle centered at C of radius y such that C does not lie on $\overline{AD}$.
- **(d)** Let Z be the intersection of line BC and the line through D which is parallel to AB.

Set $z = |DZ|$; then $\triangle CAB$ is similar to $\triangle CDZ$, so $\frac{1}{x} = \frac{y}{z}$, whence $z = xy$. □

Proposition 10. *Let $x \in \mathbb{R} \setminus \{0\}$ be constructible. Then $\frac{1}{x}$ is constructible.*

Construction. Since 1 and x are constructible, it is possible to construct line segments of length 1 and $|x|$. Without loss of generality, assume that x is positive.

- **(a)** Let A and B be points such that $|AB| = x$.
- **(b)** Let C be the point of intersection of line AB and the circle centered at A of radius 1 such that A is not on $\overline{BC}$.
- **(c)** Let D be a point of intersection of the line through A which is perpendicular to line AB and the circle centered at A of radius 1.
- **(d)** Let Z be the point of intersection of line AD and the line through C which is parallel to line BD.

Set $z = |AZ|$; then $\triangle ZAC$ is similar to $\triangle DAB$, so $\frac{z}{1} = \frac{1}{x}$, that is, $z = \frac{1}{x}$. □

A subset $F \subset \mathbb{R}$ with at least two elements is a *field* if it is closed under the operations of addition, subtraction, multiplication, and division. We have seen that the set of all constructible real numbers is a field. In particular, all rational numbers are constructible. Are there any others?

We show that the set of constructible numbers is closed under square roots; to do this, we need a couple of lemmas. Let's assume the geometric facts that the sum of angles in a triangle is 180°, and that the base angles of an equilateral triangle are equal.

Lemma 11 (Thales' Theorem)**.** *An angle inscribed in a semicircle is right.*

Proof. Consider a semicircle with center O and diameter $\overline{BC}$, and let A be an arbitrary point on the semicircle; we wish to show that $\angle BAC$ is right. Now $|OA| = |OB| = |OC|$, so $\triangle BOA$ and $\triangle COA$ are isosceles triangles. Let $\alpha = \angle OBA = \angle OAB$ and $\beta = \angle OCA = \angle OAC$; then $\angle BAC = \alpha + \beta$. Adding the angles $\triangle ABC$ we obtain

$$180^\circ = \angle OBA + \angle OCA + \angle BAC = \alpha + \beta + (\alpha + \beta) = 2(\alpha + \beta).$$

Therefore, $\angle BAC = \alpha + \beta = 90^\circ$. □

Lemma 12. *Let $\angle ACB$ be right, and let $D \in \overline{AB}$ such that $AB \perp CD$. Then $\triangle ACB \sim \triangle ADC \sim \triangle CDB$.*

Proof. Two triangles are similar if and only if they have two equal angles. Since $\angle ACB = \angle ADC = \angle CDB = 90^\circ$, $\angle DAC$ is shared by two of the triangles, and $\angle DBC$ is shared by two of the triangles, the result follows. □

Proposition 13. *Let $x \in \mathbb{R}$ be a constructible number. Then $\sqrt{|x|}$ is constructible.*

Construction. Since 1 and x are constructible, it is possible to construct line segments of length 1 and $|x|$. We may assume that x is positive.

- **(a)** Let A and B be points such that $|AB| = x$.
- **(b)** Let C be the point of intersection of line AB and the circle centered at B of radius 1 such that B is on $\overline{AC}$.
- **(c)** Let D be the midpoint of $\overline{AC}$.
- **(d)** Let Z be a point of intersection of the line through B which is perpendicular to line AB and the circle $D - A$.

Let $z = |BZ|$. Now $\angle ZBA = \angle ZBC = 90^\circ$; moreover, $\angle AZC$ is right by Thales theorem. Therefore $\triangle ZBC$ is similar to $\triangle ABZ$. Thus $\frac{z}{x} = \frac{1}{z}$, whence $z^2 = x$, so $z = \sqrt{x}$. □

10. Hippocrates Quadrature of the Lune

Proposition 14. *Any given rectangle is quadrable.*

Construction. Let $BCDE$ form a rectangle. Construct a square as follows:

- **(a)** Let F be the point of intersection of line BE and circle $E - D$ such that $E \in \overline{BF}$.
- **(b)** Let G be the midpoint of $\overline{BF}$.
- **(c)** Let H be the point of intersection of line DE and circle $G - F$ such that $E \in \overline{DH}$.
- **(d)** Let K be the point of intersection of line BE and circle $E - H$ such that $F \in \overline{EK}$.
- **(e)** Let L be the point of intersection of the line through H parallel to BE and the line through K perpendicular to BE.

Then polygon $EHLK$ is a square whose sides have length $a = |HE|$. Let $c = |BG| = |GH|$ and $b = |GE|$. Since $\triangle GEH$ is right, we have $a^2 + b^2 = c^2$. Now

$$\begin{aligned}\text{area}(BCDE) &= |BE| \times |ED| \\ &= |BE| \times |EF| \\ &= (c+b)(c-b) \\ &= c^2 - b^2 = a^2 \\ &= \text{area}(EHLK).\end{aligned}$$

□

Let P denote a plane. For $A, B, C \in P$, define the following:

- $\triangle ABC$ is the triangle whose vertices are A, B, and C.

Proposition 15. *A given triangle is quadrable.*

Construction. Let BCD form a triangle.

- **(a)** Let E be the point of intersection of line BC and the line through D which is perpendicular to BC.
- **(b)** Let F be the midpoint of $\overline{DE}$.
- **(c)** Let G be the point of intersection of the line through F which is parallel to BC and the line through B which is perpendicular to BC.
- **(d)** Let H be the point of intersection of line GF and the line through C which is perpendicular to BC.

Then $BCHG$ form a rectangle whose area is equal to the area of $\triangle BCD$. □

A *lune* is a plane region obtained by taking the complement of one disk with respect to another, where the bounding circles of the disks intersect in two points.

We now produce Hippocrates' lune. The construction uses three ingredients:

(1) the Pythagorean Theorem;
(2) an angle inscribed in a semicircle is right;
(3) the areas of two circles are to each other as the squares on their diameters.

Proposition 16. *Let A and B be points in a plane and let O be the midpoint of $\overline{AB}$. Let C be one of the points of intersection of circle $O - A$ and the line through O which is perpendicular to line AB. Let D be the midpoint of $\overline{AC}$. Let E be the point of intersection of line OD and circle $O - A$ such that $D \in \overline{OE}$. Let F be the point of intersection of line OD and circle $D - A$ such that $F \in \overline{DF}$. Then lune $AECF$ is quadrable.*

Construction. Our goal is to show that $\text{area}(\text{lune}AECF) = \text{area}(\triangle AOC)$. Note that $\angle ACB$ is a right angle, since it is inscribed in a semicircle. Triangles $\triangle AOC$ and $\triangle BOC$ are congruent by SAS; thus $|AC| = |BC|$. Apply the Pythagorean Theorem to get

$$|AB|^2 = |AC|^2 + |BC|^2 = 2|AC|^2.$$

Now

$$\frac{\text{area(semicircle } AFC)}{\text{area(semicircle } ACB)} = \frac{|AC|^2}{|AB|^2} = \frac{|AC|^2}{2|AC|^2} = \frac{1}{2}.$$

A quadrant is half of a semicircle, so clearly

$$\text{area(quadrant } ACO) = \frac{1}{2}\text{area(semicircle } ACB).$$

Thus

$$\text{area(semicircle } AFC) = \text{area(quadrant } ACO).$$

Therefore

$$\begin{aligned} \text{area}(\text{lune}AECF) &= \text{area(semicircle } AFC) - \text{area(region } AECD) \\ &= \text{area(quadrant } ACO) - \text{area(region } AECD) \\ &= \text{area}(\triangle ACO). \end{aligned}$$

□

11. Construction of Regular Polygons

A polygon is *regular* if each edge has identical length and the angles at each vertex are equal. For each positive integer n with $n \geq 3$, there is exactly one regular polygon with n edges, up to similarity; is is called a regular n-gon.

Let us first determine the angles in a regular n-gon. It can be inscribed in a circle, and so has a specific center. Divide the n-gon into n isosceles triangles, each with adjacent vertices on the n-gon, with the third vertex being the center. Note that the base angles bisect the angles of the n-gon. Now the sum of the angles of the triangles which come together at the center is $360°$. Thus the base angles add to $180°n - 360°$. There are $2n$ congruent base angles, so each has size

$$\frac{180°n - 360°}{2n} = 90°\left(1 - \frac{2}{n}\right).$$

The angles of the n-gon consist of two base angles, so each angle of the n-gon is

$$180°\left(1 - \frac{2}{n}\right).$$

We may canonically inscribe a regular polygon with n edges in the unit circle of the cartesian plane; its set of vertices is

$$\{(\cos\alpha, \sin\alpha) \in \mathbb{R}^2 \mid \alpha = \frac{2\pi k}{n} \text{ for } k = 0, 1, \ldots, n-1\}.$$

This is a convenient way for us to view a regular polygon: for example, the length of one side is the distance from $(1, 0)$ to $(\cos\alpha, \sin\alpha)$, where $\alpha = \frac{2\pi}{n}$. By the distance formula,

$$\begin{aligned} \text{length(edge)} &= \sqrt{(\cos\alpha - 1)^2 + (\sin\alpha - 0)^2} \\ &= \sqrt{\cos^2\alpha + \sin^2\alpha + 1 - 2\cos\alpha} \\ &= \sqrt{2 - 2\cos\alpha}. \end{aligned}$$

The ancient Greeks, however, had no coordinate system; they attempted to construct regular polygons using straight-edge and compass.

If a line segment of length r is given, we see that constructibility of a regular n-gon is equivalent to the constructibility of the real number $r\cos\frac{360°}{n}$. We reserve the right to use this existence criterion later, but we begin with actual constructions. All of our constructions proceed from a line segment of length r, and are inscribed in a circle of radius r.

Let O and A be given point with $|OA| = r$. If we can construct a point Z such to $\angle AOZ = \frac{360°}{n}$, then we can complete the construction of the other vertices by intersecting circles centered at a previously constructed vertex of radius $|AZ|$ with circle $O - A$. Thus, it suffices to construct such a point Z.

Proposition 17. *A regular triangle is constructible from* $\{O, A\}$.

Proof. We are given O and A.

- **(a)** Let B be the point of intersection of line OA and circle $O - A$ which is not A.
- **(b)** Let C be the midpoint of $\overline{OB}$.
- **(c)** Let Z be a point of intersection of the line through C which is perpendicular to line OA and circle $O - A$ such that $\angle AOZ \leq 180°$.

Now $\angle AOZ = 120°$. □

Proposition 18. *A square is constructible from* $\{O, A\}$.

Proof. We are given O and A.

- **(a)** Let Z be the point of intersection of the line through O which is perpendicular to line OA and circle $O - A$ such that $\angle AOZ \leq 180°$.

Now $\angle AOZ = 90°$. □

Proposition 19. *If a regular* n*-gon is constructible, then so is a regular* $2n$*-gon.*

Construction. We are given O and A.

- **(a)** Let B be a point on the circle $O - A$ such that $\angle AOB = \frac{360°}{n}$.
- **(b)** Let C be the midpoint of $\overline{AB}$.
- **(c)** Let Z be the intersection of line OC and circle $O - A$ such that Z lies on $\overline{AB}$.

Now $\angle AOZ = \frac{360°}{2n}$. □

Thus we may construct regular triangles, quadrilaterals, and hexagons. We would like to know if a regular pentagon is constructible. Investigating this brings us to the world of the golden ratio.

12. Problems

For each construction, provide a drawing produced with an actual straightedge and compass, together with a list of steps sufficient to reproduce the drawing (as in the propositions of the text). If you apply the propositions to construct a midpoint or perpendicular, use a marked ruler or protractor to obtain a more accurate picture.

Problem 1. For each subset Q of a plane P, find all points that are immediately constructible from Q.

- **(a)** Q consists of two points
- **(b)** Q consists of the vertices of an equilateral triangle
- **(c)** Q consists of the vertices of an isosceles triangle

Problem 2. Reproduce the drawings which correspond to the construction instructions for the following propositions.

- **(a)** Proposition 1 (midpoints)
- **(b)** Proposition 3 (perpendiculars)
- **(c)** Proposition 5 (transference of distance)
- **(d)** Proposition 10 (products of constructible lengths)
- **(e)** Proposition 11 (quotients of constructible lengths)
- **(f)** Proposition 12 (square roots of constructible lengths)
- **(g)** Proposition 13 (quadrature of a rectangle)
- **(h)** Proposition 14 (quadrature of a triangle)
- **(i)** Proposition 15 (quadrature of a lune)

Problem 3. Given circle $A - B$, construct an equilateral triangle inscribed in the circle with one vertex at B.

Problem 4. Given circle $A - B$, construct a regular hexagon inscribed in the circle with one vertex at B.

Problem 5. Given three noncollinear points, construct the center of the unique circle which contains the three points.

Problem 6. Given two points, construct a line segment of length $\sqrt{3}$.

Problem 7. Given two points, construct a line segment of length $\sqrt{2}$.

Problem 8. Given two points, construct an angle of $45°$.

Problem 9. Given two points, construct an angle of $75°$.

Problem 10. Given a circle, construct a concentric circle with quadruple the area.

Problem 11. Given a circle, construct a concentric circle with triple the area.

Problem 12. Given a circle, construct a concentric circle with double the area.

CHAPTER III

The Golden Section

1. The Golden Section

Let A and B be points in a plane. A *section* of $\overline{AB}$ is a point C in the interior of $\overline{AB}$. Consider the case where $|AC| \geq |CB|$; here are various ratios of the lengths of the segments that can be explored, for example $\frac{|AB|}{|AC|}$ and $\frac{|AC|}{|BC|}$.

A *golden section* of $\overline{AB}$ is section C of $\overline{AB}$ which satisfies

$$\frac{|AB|}{|AC|} = \frac{|AC|}{|BC|}.$$

In this case, the common value of these fractions is known as the *golden ratio*; this clearly does not depend on the length of $\overline{AB}$. Thus the golden ratio is a specific, well-defined number which we denote by the Greek letter φ.

Let $x = |AB|$, $y = |AC|$, and $z = |CB|$. In the case of a golden section, $\frac{x}{y} = \frac{y}{z}$, so that $xz = y^2$. Moreover, $x = y + z$, and substituting this into the previous equation and rearranging, we obtain

$$y^2 - zy - z^2 = 0.$$

Then the quadratic formula gives

$$y = \frac{z \pm \sqrt{z^2 + 4z^2}}{2} = z\frac{1 \pm \sqrt{5}}{2}.$$

Since $\sqrt{5} > 1$ and y cannot be negative, one of these solutions is spurious. In the ratio $\frac{y}{z}$, the z's cancel and we obtain

$$\boxed{\varphi = \frac{1 + \sqrt{5}}{2}.}$$

What percentage of a given line segment is a golden section?

$$\frac{y}{x} = \frac{1}{\phi} = \frac{2}{\sqrt{5}+1} = \frac{2(\sqrt{5}-1)}{5-1} = \frac{\sqrt{5}-1}{2} \approx 0.62;$$

also,

$$\frac{z}{x} = \frac{x-y}{x} = 1 - \frac{y}{x} = 1 - \frac{\sqrt{5}-1}{2} = \frac{3-\sqrt{5}}{2} \approx 0.38.$$

2. Recreational Appearances of the Golden Ratio

We see that the golden ratio is the positive solution to the polynomial equation $x^2 - x - 1$. In particular,

$$\varphi^2 = \varphi + 1.$$

Moreover, dividing this equation by φ and subtracting 1 from both sides yields

$$\frac{1}{\varphi} = \varphi - 1.$$

So here we have a number whose square is obtained by adding 1, and whose inverse is obtained by subtracting 1.

Consider the continued square root

$$\sqrt{1 + \sqrt{1 + \sqrt{1 + \ldots}}}.$$

Assuming that this pattern is meaningful and represents a number, let x be that number. Then clearly $x > 0$. Squaring $x = \sqrt{1 + \sqrt{1 + \sqrt{1 + \ldots}}}$ yields

$$x^2 = 1 + \sqrt{1 + \sqrt{1 + \ldots}} = 1 + x.$$

Thus x satisfies $x^2 - x - 1 = 0$, so $x = \varphi$.

Consider the continued fraction

$$1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \ldots}}}.$$

Again assume that this pattern represents some number x; we see that

$$x = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \ldots}}} = 1 + \frac{1}{x}.$$

Multiplying both sides by x gives $x^2 = x + 1$. Thus again we see that $x = \varphi$.

Let's attempt to make this example more precise by restating it using the language of sequences. We wish to construct a (hopefully convergent) sequence $(a_n)_{n \in \mathbb{N}}$ such that each a_n is a fraction representing an approximation of the above continued fraction, with increasing accuracy, so that the limit would be the inescapable meaning of the above continued fraction. Let's begin with $1 + \frac{1}{2}$, and at each stage replace 2 with $1 + \frac{1}{2}$. We obtain

$$\begin{aligned}
a_1 &= 1 + \frac{1}{2} && = \frac{3}{2} \\
a_2 &= 1 + \frac{1}{1 + \frac{1}{2}} && = \frac{5}{3} \\
a_3 &= 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{2}}} && = \frac{8}{5} \\
a_4 &= 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{2}}}} && = \frac{13}{8}
\end{aligned}$$

and so forth. Can you guess the value of a_5? Does this relate to anything else you have previously seen?

3. Construction of the Golden Section

We now describe how to construct a golden section of a given line segment. The idea is to construct a right triangle such that one leg is twice as long as the other, so that by the Pythagorean theorem, the hypotenuse will contain a square root of 5.

Proposition 1. *A golden section is constructible.*

Construction. We are given line segment $\overline{AB}$; we construct a point Z between A and B such that $\frac{|AB|}{|AZ|} = \frac{|AZ|}{|ZB|}$, or equivalently, such that $|AZ| = \frac{\sqrt{5}-1}{2}$.

- **(a)** Let D be the point of intersection of line AB and circle $A - B$ which is not B.
- **(b)** Let E be the midpoint of $\overline{DA}$.
- **(c)** Let F be the point of intersection of circle $A - B$ and the line through A perpendicular to A.
- **(d)** Let Z be the point of intersection of line AB and circle $E - F$ which lies on $\overline{AB}$.

To see that this succeeds, scale our situation so that $|AB| = 1$. Then $|DA| = 1$, so $|EA| = \frac{1}{2}$. Also, $|FA| = 1$, so by the Pythagorean Theorem, $|EF| = |EZ| = \sqrt{1^2 + \frac{1}{4}} = \frac{\sqrt{5}}{2}$. Thus $|AZ| = |EZ| - |EA| = \frac{\sqrt{5}-1}{2}$. □

4. The Golden Rectangle

Consider a rectangle $ABDC$ such that sides $\overline{AB}$ and $\overline{CD}$ are the longer sides, with length x, and that sides $\overline{AC}$ and $\overline{BD}$ are shorter, with length y. Let E and F lie on $\overline{AB}$ and $\overline{CD}$, respectively, so that $AEFC$ is a square. We call rectangle $ABDC$ a *golden rectangle* if rectangle $ABDC$ is similar to rectangle $FEBD$.

Suppose that rectangle $ABCD$ is golden, and let $z = |EB|$; then $x = y + z$. By similarity, we have $\frac{x}{y} = \frac{y}{z}$, which leads to $y^2 - zy - z^2 = 0$. We see that E and F cut $\overline{AB}$ and $\overline{BD}$ in a golden section, and $\frac{x}{y} = \varphi$. Thus a golden rectangle is constructible as a rectangle build on a golden section.

Proposition 2. *A golden rectangle is constructible.*

Construction. We are given point A and B which form one side of the rectangle.

- **(a)** Let C be a golden section of $\overline{AB}$, with longer side $\overline{AC}$.
- **(b)** Let D be the point of intersection of circle $A - C$ and the line through A which is perpendicular to AB.
- **(c)** Let E be the point of intersection of the line through B which is perpendicular to B, and the line through D which is parallel to AB.

Then $ABED$ is a golden rectangle. □

5. The Golden Triangle

Consider an isosceles triangle $\triangle ABC$, where $\angle ABC = \angle ACB$. Let D be the point of intersection of line AC and a bisector of angle $\angle ABC$. We call $\triangle ABC$ a *golden triangle* if $\triangle ABC$ is similar to $\triangle BDC$.

Suppose that $\triangle ABC$ is golden, and let $x = |AB| = |AC|$ and $y = |BC|$. Then $\triangle BDC$ is isosceles, and $|BD| = |BC| = y$. Also $\angle BAC = \angle ABD$, so $\triangle DAB$ is also an isosceles triangle, and $|AD| = |BD| = y$. Let $z = |DC|$; then $x = y + z$. By similarity, we have $\frac{x}{y} = \frac{y}{z}$; therefore, as before, $\frac{x}{y} = \varphi$.

We may compute the angles of a golden triangle as follows. Let $\alpha = \angle BAC$ and $\beta = \angle ABC = \angle ACB$, so that $\beta = 2\alpha$. Then $5\alpha = 180^\circ$, so $\alpha = 36^\circ$ and $\beta = 72^\circ$.

This allows us to compute $\cos 72^\circ$; construct a right triangle $\triangle AEB$ by letting E be the midpoint of $\overline{BC}$. Then $|BE| = \frac{y}{2}$, so

$$\cos\beta = \frac{y}{2x} = \frac{1}{2\varphi}.$$

Since $\frac{1}{\varphi} = \varphi - 1$, conclude that

$$\boxed{\cos 72^\circ = \frac{-1+\sqrt{5}}{4}.}$$

This fact will help us in the construction of a golden triangle.

Proposition 3. *A golden triangle is constructible.*

Construction. We are given points A and D; we construct points B and C so that $\triangle ABC$ is golden, with base $\overline{AB}$.

- **(a)** Let B be a golden section of $\overline{AD}$, we longer side $\overline{AB}$.
- **(b)** Let C be the point of intersection of the circle $A - D$ and the line through B which is perpendicular to AD.

□

6. Construction of a Regular Pentagon

We are given two points O and A, and we wish to construct a regular pentagon inscribed in the circle $O - A$ such that A is one of the vertices. If Z is a vertex adjacent to A, then $\angle AOZ = \frac{360^\circ}{5} = 72^\circ$. Thus if we can construct on $\overline{OA}$ a section Y such that $|OY| = \cos 72^\circ = \frac{-1+\sqrt{5}}{4}$, we will be well on our way to construction of the regular pentagon. We have seen that this is possible; we repeat the construction here.

Proposition 4. *A regular pentagon is constructible.*

Construction. We are given point O and A with $|OA| = r$. For simplicity and without loss of generality, assume that $r = 1$.

- **(a)** Let B be the point of intersection of line OA and circle $O - A$ which is not A.
- **(b)** Let C be the midpoint of $\overline{BO}$.
- **(c)** Let D be a point of intersection of the line through O which is perpendicular to OA, and the circle $O - A$.
- **(d)** Let E be the point of intersection of line OA and circle $C - D$.
- **(e)** Let F be the midpoint of $\overline{OE}$.
- **(f)** Let Z be the point of intersection of circle $O - A$ and the line through F which is perpendicular to OA.

Then $\angle AOZ = 72^\circ$, so that $\overline{AZ}$ is the side of a regular pentagon inscribed in circle $O - A$. The other sides are now attainable. □

7. The Golden Pentagram

The *diagonals* of a regular pentagon are the line segments between non-adjacent edges. There are five such diagonals; there union is known as the *golden pentagram*. This star-shaped figure was used as the logo of the Pythagorean brotherhood.

Let A, B, C, D, and E be the vertices of a regular pentagon, labeled in counterclockwise order. Label the points of intersection of the diagonals as follows: $F \in \overline{AC} \cap \overline{BE}$, $G \in \overline{BD} \cap \overline{CA}$, $H \in \overline{CE} \cap \overline{DB}$, $I \in \overline{DA} \cap \overline{EC}$, and $J \in \overline{EB} \cap \overline{AD}$.

We wish to show that $\triangle ACD$ is a golden triangle, and that polygon $FGHIJ$ is another regular pentagon.

Let $\alpha = \angle CAD$, $\beta = \angle ACD$, $\gamma = \angle BAC$, and $\delta = \angle BAE$.

By the formula for the angles of a regular polygon, we have

$$\delta = 180^\circ(1 - \frac{2}{n}) = 108^\circ.$$

Since pentagon $ABCDE$ is regular, the Side-Angle-Side Theorem implies that

$$\triangle ABC \cong \triangle BCD \cong \triangle CDE \cong \triangle DEA \cong \triangle EAB,$$

where the symbol $\cong$ means "is congruent to"; moreover, these are all isosceles triangles. Thus $\gamma = \angle ABE$. This shows that $\triangle FAB$ is similar to $\triangle ABE$, which is isosceles; thus $\angle AFB = \delta$, so $2\gamma + \delta = 180^\circ$, which gives

$$\gamma = \frac{180^\circ - \delta}{2} = 36^\circ.$$

Similarly, we have $\gamma = \angle DAE$, so $\angle BAE = \delta = 2\gamma + \alpha$, so

$$\alpha = \delta - 2\gamma = \gamma = 36^\circ.$$

Now $\overline{AC} = \overline{AD}$ because $\triangle BAC \cong \triangle EAD$, so $\triangle ACD$ is isosceles. Thus $\alpha + 2\beta = 180^\circ$, so

$$\beta = \frac{180^\circ - \alpha}{2} = 72^\circ.$$

Thus $\triangle CAD$ is a golden triangle.

Similarly, one finds other golden triangles in this diagram; we see that

$$\triangle ACD \cong \triangle BDC \cong \triangle CEA \cong \triangle DAB \cong \triangle EBC.$$

We also see that

$$\triangle ABG \cong \triangle BCH \cong \triangle CDI \cong \triangle DEJ \cong \triangle EAF,$$

and

$$\triangle AFJ \cong \triangle BGF \cong \triangle CHG \cong \triangle DIH \cong \triangle EJI$$

are sets of congruent golden triangles. From this, polygon $FGHIJ$ is a regular pentagon.

8. Incommensurability

Let A, B, C, and D be points in a plane. We say that $\overline{AB}$ and $\overline{CD}$ are *commensurable* if there exists a line segment $\overline{EF}$ and positive integers m and n such that

$$|AB| = m\overline{EF} \quad \text{and} \quad |CD| = n\overline{EF};$$

thus $\frac{|AB|}{|CD|} = \frac{m}{n}$. The Pythagoreans assumed in their proofs that any two line segments as commensurable. Suppose that $\overline{CD} = 1$; then this assumption amounts to

$$|AB| = \frac{m}{n} \in \mathbb{Q},$$

that is, the length of any line segment is a rational number.

Thus for the Pythagoreans, it must have been quite a shock to realize that not all constructible numbers are rational. This may have been discovered during contemplation of the golden pentagram, and follows.

Continue notation from the previous section. Since $\triangle ACD$ is golden, we have $|AC|/|CD| = \varphi$. Now $|CD| = |CI|$ and $\triangle CDI$ is golden, so $|CD|/|DI| = \varphi$. But then $\triangle DIH$ is golden, so $|IH|/|DH| = \varphi$. At this point, we notice that $\overline{HI}$ is an edge of the regular pentagon $FGHIJ$, and the diagonals of this pentagon have length $|DH|$. If we inscribe another pentagram in this pentagon, we see that this chain of equalities will continue forever.

Now if all line segments are commensurable, there exists a line segment $\overline{MN}$ such that $|AC|$ and $|CD|$ are in integer multiples of $|MN|$. Now $|AI| = |CD|$, so $|AI|$ is also an integer multiple of $|MN|$. This shows that

$$|DI| = |AD| - |AI|$$

is also a multiple of $|mn|$. Repeating this argument shows that $|HI|$ is an integer multiple of $|MN|$, and this continues into the smaller pentagon.

We can continue this process, getting smaller and smaller pentagons with smaller and smaller edges, but each edge will be an integral multiple of some line segment $\overline{MN}$ of fixed length. Perhaps it was this contradiction which first demonstrated the existence of irrational numbers.

9. Regular Solids

Recall that a *polygon* is a plane figure bounded by line segments. A plane region is *convex* if, given any to points in the interior, the line segment between these points is contained in the interior.

Recall that a *polyhedron* is a space figure bounded by polygons. The bounding polygons are called *faces*, the bounding line segments of these polygons are called *edges*, and the endpoints of these line segments are called *vertices*. Again, a space region is *convex* if, given any to points in the interior, the line segment between these points is contained in the interior.

A polyhedron is *regular* if

(a) it is convex;
(b) its faces of congruent regular polygons;
(c) its vertices have the same number of attached edges.

Regular polyhedra are also known as regular solids, or as Platonic solids. We wish to classify the regular solids.

First, we decide what the possibilities are, and then we describe the construction of each possibility.

The key to deciding the possibilities is to realize that if multiple faces come together are a vertex, there must be at least three faces, and the sum of the angles which come together must be less than $360°$.

The following chart indicates the possibilities. The first column represents the number of sides of the polygonal faces. By the formula angle $= 180°(1 - \frac{2}{n})$, we compute the internal angles of a regular n-gon. Then we see how many faces can come together are a vertex.

Sides	Angle/Vertex	Faces/Vertex	Total Angle	Possible?
3	$60°$	3	$180°$	Yes
3	$60°$	4	$240°$	Yes
3	$60°$	5	$300°$	Yes
3	$60°$	≥ 6	$\geq 360°$	No
4	$90°$	3	$270°$	Yes
4	$90°$	≥ 4	$\geq 360°$	No
5	$108°$	3	$324°$	Yes
5	$108°$	≥ 4	$\geq 432°$	No
≥ 6	$\geq 120°$	≥ 3	$\geq 360°$	No

So we have five possibilities; there are at most five regular solids (up to similarity). Next we demonstrate that each of the five possibilities exist.

10. Construction of the Regular Solids

We wish to construct each regular solid using Euclidean tools (even through we analyze the construction using analytic geometry). It suffices to construct the vertices in $\mathbb{R}^3$ from the set $Q = \{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\} \subset \mathbb{R}^3$.

First start with 3 squares coming together at a vertex. This will form a a solid with six sides which we may call a Hexahedron, but is usually known as a cube, The cube is easily constructed from the set Q; for example, $(1,1,0)$ is the intersection of a line on the xy-plane perpendicular to the x-axis through $(1,0,0)$, and a line on the xy-plane perpendicular to the y-axis through $(0,1,0)$. The complete vertex set is

$$\text{Cube} = \{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}.$$

Next construct a regular solid with 3 equilateral triangles coming together at each vertex; this solid will have 4 faces, and is thus known as a regular tetrahedron. We see tetrahedra embedded in the cube by drawing line segments diagonally across the faces; this will create two sets of vertices of regular tetrahedra. One of these sets is

$$\text{Tetrahedron} = \{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\}.$$

This will produce regular faces if all of the edges have the same length. Computation shows that indeed, the edges have length $\sqrt{2}$.

Now we wish to produce a regular solid with 4 equilateral triangles coming together at each vertex; this solid will have 8 faces, and so it is known as a regular octahedron. To construct a regular octahedron, take its vertices to be the set of centers of the faces of the cube; this will give 6 vertices; take the cube to have sides of length two to simplify the situation, and find that

$$\text{Octahedron} = \{(1,1,0),(1,0,1),(0,1,1),(1,1,2),(1,2,1),(2,1,1)\}.$$

The lengths of the edges of this solid are also $\sqrt{2}$.

We note that if we take the centers of the faces of an octahedron as vertices for a solid, we obtain a cube; that is, the 8 vertices of the cube correspond to the 8 faces of the octahedron, just as the 6 vertices of the octahedron correspond to the 6 faces of the cube. We say that the cube and the octahedron are *dual* polyhedra. Note that the dual of the tetrahedron is another tetrahedron; it is *self-dual*.

Next we construct a solid with five triangular faces coming together at each vertex, which has twenty faces and as such is known as an icosahedron. To do this, embed three golden rectangles with sides of length 1 and φ in $\mathbb{R}^3$ on the coordinate planes so that the center of each rectangle is the origin.

Let $\alpha = \frac{1}{2}$ and let $\beta = \frac{1+\sqrt{5}}{4} = \frac{\varphi}{2}$. Set

$$\text{Icosahedron} = \{(0,\pm\alpha,\pm\beta),(\pm\alpha,\pm\beta,0),(\pm\beta,0,\pm\alpha)\},$$

this set contains 12 points, and produces a solid with 20 triangular faces. For example, one of the faces has vertices $A = (\beta,\alpha,0)$, $B = (\beta,-\alpha,0)$, and $C = (\alpha,0,\beta)\}$. That $|AB| = 1$ is clear, and that $|AC| = |BC|$ is also clear. To see

that this is an equilateral triangle, we compute

$$\begin{aligned}|AC| &= \sqrt{(\alpha-\beta)^2+(0-\alpha)^2+(\beta-0)^2}\\ &= \sqrt{2\alpha^2+2\beta^2-2\alpha\beta}\\ &= \sqrt{\frac{1}{2}+\frac{\varphi^2}{2}-\frac{\varphi}{2}}\\ &= \sqrt{\frac{1+(\varphi+1)-\varphi}{2}}\\ &= 1.\end{aligned}$$

Thus indeed, we have constructed a regular triangle, so we have a regular icosahedron.

Finally, we consider the case of three regular pentagons coming together at a vertex; this produces a polyhedron with twelve faces known as a dodecahedron. We can obtain this as the dual of the icosahedron; that is, let the vertex set be the set of centers of the faces of a regular icosahedron.

We investigate this vertex set. The center of an equilateral triangle in space is obtained by averaging the coordinates of the vertices; that is, the center of the equilateral triangle $\triangle A_1A_2A_3$, where $A_i=(x_i,y_i,z_i)$, is

$$\left(\frac{x_1+x_2+x_3}{3},\frac{y_1+y_2+y_3}{3},\frac{z_1+z_2+z_3}{3}\right).$$

In our case, we obtain two types of triangles: those who share a side with one of the golden rectangles, and those whose vertices come from three different golden rectangles. There are twelve of the former and eight of the latter.

The first twelve are easy to see: there are six sides of length 1 on the rectangles, and two triangles sharing each such side. The eight others are obtained by noticing that only certain combinations are possible. Here is a complete list, with all coordinates multiplied by three, $\gamma=\alpha+2\beta$, and $s_1,s_2,s_3\in\{\pm1\}$.

$$\begin{aligned}\text{Dodecahedron} = \{&(\pm\gamma,0,\pm\beta),(\pm\beta,\pm\gamma,0),(0,\pm\beta,\pm\gamma),\\ &(s_1\beta,s_2\alpha,0),(0,s_2\beta,s_3\alpha),(s_1\alpha,0,s_3\beta)\}.\end{aligned}$$

11. Problems

Problem 1. Consider the sequence (a_n) of real numbers defined by

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

Assuming that (a_n) converges, find $\lim_{n\to\infty} a_n$. To prove that (a_n) converges, show that (a_n) is bounded and increasing.

Problem 2. Consider the sequence (a_n) of real numbers defined by

$$a_1 = 1 \quad \text{and} \quad a_{n+1} = \frac{1}{1 + a_n}.$$

Assuming that (a_n) converges, find $\lim_{n\to\infty} a_n$.

Problem 3. Consider a pyramid with four triangular sides and a square base. Let h be the height of the pyramid. Let s be the height of a triangular side, let a be half the length of its base, so that the area of the triangular side is sa. Show that if $h^2 = sa$, then the slope of the pyramid, $\frac{s}{a}$, is equal to the Golden Ratio.

Problem 4. Consider the regular solids inscribed in a unit sphere.

- **(a)** Find the lengths of the line segments for each solid.
- **(b)** Find the area of a face of each solid.
- **(c)** Find the angle between the faces of each solid.
- **(d)** Find the radius of the inscribed sphere of each solid.
- **(e)** Find the volume of each solid.

CHAPTER IV

The Euclidean Algorithm

1. Induction and the Well-Ordering Principle

First we establish a few properties of the integers which we need in order to develop the Euclidean algorithm. We start with the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$, and accept the Peano Axioms as a characterization of $\mathbb{N}$. The primary axiom is stated below.

Proposition 1. Peano's Axiom
Let $S \subset \mathbb{N}$. If

(a) $1 \in S$, *and*
(b) $n \in S \Rightarrow n + 1 \in S$,

then $S = \mathbb{N}$.

From this, the Well-Ordering Principle follows.

Proposition 2. Well-Ordering Principle
Let $X \subset \mathbb{N}$ be a nonempty set of positive integers. Then X contains a smallest, element; that is, there exists $a \in X$ such that for every $x \in X$, $a \leq x$.

Proof. Let $X \subset \mathbb{N}$ and assume that X has no smallest element; we show that $X = \varnothing$. Let

$$S = \{n \in \mathbb{N} \mid n < x \text{ for every } x \in X\}.$$

Clearly $S \cap X = \varnothing$; if we show that $S = \mathbb{N}$, then $X = \varnothing$.

Since 1 is less than every natural number, 1 is less than every natural number in X. Thus $1 \in X$.

Suppose that $n \in S$. Then $n < x$ for every $x \in X$, so $n + 1 \leq x$ for every $x \in X$. If $n+1$ were in X, it would be the smallest element of X; since X has no smallest element, $n + 1 \notin x$; thus $n + 1 \neq x$ for every $x \in X$, whence $n + 1 < x$ for every $x \in X$. It follows that $n + 1 \in S$, and by Peano's Axiom, $S = \mathbb{N}$. $\square$

2. Division Algorithm

Proposition 3. Division Algorithm for Integers
Let $m, n \in \mathbb{Z}$. There exist unique integers $q, r \in \mathbb{Z}$ such that

$$n = qm + r \qquad \text{and} \qquad 0 \leq r < m.$$

Proof. Let $X = \{z \in \mathbb{Z} \mid z = n - km \text{ for some } k \in \mathbb{Z}\}$. The subset of X consisting of nonnegative integers is a subset of $\mathbb{N}$, and by the Well-Ordering Principle, contains a smallest member, say r. That is, $r = n - qm$ for some $q \in \mathbb{Z}$, so $n = qm + r$. We know $0 \leq r$. Also, $r < m$, for otherwise, $r - m$ is positive, less than r, and in X.

For uniqueness, assume $n = q_1m + r_1$ and $n = q_2m + r_2$, where $q_1, r_1, q_2, r_2 \in \mathbb{Z}$, $0 \leq r_1 < m$, and $0 \leq r_2 < m$. Then $m(q_1 - q_2) = r_1 - r_2$; also $-m < r_1 - r_2 < m$. Since $m \mid (r_1 - r_2)$, we must have $r_1 - r_2 = 0$. Thus $r_1 = r_2$, which forces $q_1 = q_2$. □

Definition 4. Let $m, n \in \mathbb{Z}$. We say that m *divides* n, and write $m \mid n$, if there exists an integer k such that $n = km$.

Definition 5. Let $m, n \in \mathbb{Z}$. A *greatest common divisor* of m and n, denoted $\gcd(m, n)$, is a positive integer d such that

(1) $d \mid m$ and $d \mid n$;
(2) If $e \mid m$ and $e \mid n$, then $e \mid d$.

Proposition 6. *Let $m, n \in \mathbb{Z}$. Then there exists a unique $d \in \mathbb{Z}$ such that $d = \gcd(m, n)$, and there exist integers $x, y \in \mathbb{Z}$ such that*

$$d = xm + yn.$$

Proof. Let $X = \{z \in \mathbb{Z} \mid z = xm + yn \text{ for some } x, y \in \mathbb{Z}\}$. Then the subset of X consisting of positive integers contains a smallest member, say d, where $d = xm + yn$ for some $x, y \in \mathbb{Z}$.

Now $m = qd + r$ for some $q, r \in \mathbb{Z}$ with $0 \leq r < d$. Then $m = q(xm + yn) + r$, so $r = (1 - qxm)m + (qy)n \in X$. Since $r < d$ and d is the smallest positive integer in X, we have $r = 0$. Thus $d \mid m$. Similarly, $d \mid n$.

If $e \mid m$ and $e \mid n$, then $m = ke$ and $n = le$ for some $k, l \in \mathbb{Z}$. Then $d = xke + yle = (xk + yl)e$. Therefore $e \mid d$. This shows that $d = \gcd(m, n)$.

For uniqueness of a greatest common divisor, suppose that e also satisfies the conditions of a gcd. Then $d \mid e$ and $e \mid d$. Thus $d = ie$ and $e = jd$ for some $i, j \in \mathbb{Z}$. Then $d = ijd$, so $ij = 1$. Since i and j are integers, then $i = \pm 1$. Since d and e are both positive, we must have $i = 1$. Thus $d = e$. □

Fact 7. *Let $m, n \in \mathbb{Z}$ and suppose that there exist integers $x, y \in \mathbb{Z}$ such that $xm + yn = 1$. Show that $\gcd(m, n) = 1$.*

Fact 8. *Let $m, n \in \mathbb{N}$ and suppose that $m \mid n$. Show that $\gcd(m, n) = m$.*

3. Euclidean Algorithm

There is an efficient effective procedure for finding the greatest common divisor of two integers. It is based on the following proposition.

Proposition 9. *Let $m, n \in \mathbb{Z}$, and let $q, r \in \mathbb{Z}$ be the unique integers such that $n = qm + r$ and $0 \leq r < m$. Then $\gcd(n, m) = \gcd(m, r)$.*

Proof. Let $d_1 = \gcd(n, m)$ and $d_2 = \gcd(m, r)$. Since "divides" is a partial order on the positive integers, it suffices to show that $d_1 \mid d_2$ and $d_2 \mid d_1$.

By definition of common divisor, we have integers $w, x, y, z \in \mathbb{Z}$ such that $d_1 w = n$, $d_1 x = m$, $d_2 y = m$, and $d_2 z = r$.

Then $d_1 w = q d_1 x + r$, so $r = d_1(w - qx)$, and $d_1 \mid r$. Also $d_1 \mid m$, so $d_1 \mid d_2$ by definition of gcd.

On the other hand, $n = q d_2 y + d_2 z = d_2(qy + z)$, so $d_2 \mid n$. Also $d_2 \mid m$, so $d_2 \mid d_1$ by definition of gcd. □

Now let $m, n \in \mathbb{Z}$ be arbitrary integers, and write $n = mq + r$, where $0 \leq r < m$. Let $r_0 = n$, $r_1 = m$, $r_2 = r$, and $q_1 = q$. Then the equation becomes $r_0 = r_1 q_1 + r_2$. Repeat the process by writing $m = rq_2 + r_3$, which is the same as $r_1 = r_2 q_2 + r_3$, with $0 \leq r_3 < r_2$. Continue in this manner, so in the i^{th} stage, we have $r_{i-1} = r_i q_i + r_{i+1}$, with $0 \leq r_{i+1} < r_i$. Since r_i keeps getting smaller, it must eventually reach zero.

Let k be the smallest integer such that $r_{k+1} = 0$. By the above proposition and induction,

$$\gcd(n, m) = \gcd(m, r) = \cdots = \gcd(r_{k-1}, r_k).$$

But $r_{k-1} = r_k q_k + r_{k+1} = r_k q_k$. Thus $r_k \mid r_{k-1}$, so $\gcd(r_{k-1}, r_k) = r_k$. Therefore $\gcd(n, m) = r_k$. This process for finding the gcd is known as the *Euclidean Algorithm*.

In order to find the unique integers x and y such that $xm + yn = \gcd(m, n)$, use the equations derived above and work backward. Start with $r_k = r_{k-2} - r_{k-1}q_{k-1}$. Substitute the previous equation $r_{k-1} = r_{k-3} - r_{k-2}q_{k-2}$ into this one to obtain

$$r_k = r_{k-2} - (r_{k-3} - r_{k-2}q_{k-2})q_{k-1}) = r_{k-2}(q_{k-2}q_{k-1} + 1) - r_{k-3}q_{k-1}.$$

Continuing in this way until you arrive back at the beginning.

For example, let $n = 210$ and $m = 165$. Work forward to find the gcd:

- $210 = 165 \cdot 1 + 45$;
- $165 = 45 \cdot 3 + 30$;
- $45 = 30 \cdot 1 + 15$;
- $30 = 15 \cdot 2 + 0$.

Therefore, $\gcd(210, 165) = 15$. Now work backwards to find the coefficients:

- $15 = 45 - 30 \cdot 1$;
- $15 = 45 - (165 - 45 \cdot 3) = 45 \cdot 4 - 165$;
- $15 = (210 - 165) \cdot 4 - 165 = 210 \cdot 4 - 165 \cdot 5$.

Therefore, $15 = 210 \cdot 4 + 165 \cdot (-5)$.

Let's briefly analyze the inductive process of "working backwards".

At each stage, let m denote the smaller number and let n denote the larger number. Always attach x to m and y to n, to get $d = xm + yn$, where $d = \gcd(m, n)$. Now at the very end, the remainder is zero, so

$$n = mq + 0.$$

Thus $m = \gcd(n, m)$, that is, $d = m$. Writing d as a linear combination at this stage, we have

$$d = (1)m + (0)nm$$

so $x = 1$ and $y = 0$.

Now we want to lift this to a previous equation of the form $n = mq + r$. Assume, by way of induction, that we have already lifted it to the next equation; that is, we have $n' = m'q' + r'$, where $n' = m$, $m' = r$, and we can express d as a linear combination of m' and n', like this:

$$d = x'm' + y'n'.$$

Then $d = x'r + y'm$. Substitute in $r = n - mq$ to express d as a linear combination of m and n; you get $d = x'(n - mq) + y'm = (y' - x'q)m + x'n$. Set $x = y' - x'q$ and $y = x'$ to obtain $d = xm + yn$.

4. Fundamental Theorem of Arithmetic

Definition 10. An integer $p \geq 2$, is called *prime* if

$$a \mid p \Rightarrow a = 1 \text{ or } a = p, \quad \text{where } a \in \mathbb{N}.$$

An integer $n \geq 2$ is called *composite* if it is not prime.

Proposition 11. *Let $p \in \mathbb{Z}$, $p \geq 2$. Then p is prime if and only if*

$$p \mid ab \Rightarrow p \mid a \text{ or } p \mid b, \quad \text{where } a, b \in \mathbb{N}.$$

Proof.
($\Rightarrow$) Given that $a \mid p \Rightarrow a = 1$ or $a = p$, suppose that $p \mid ab$. Then there exists $k \in \mathbb{N}$ such that $kp = ab$. Suppose that p does not divide a; then $\gcd(a, p) = 1$. Thus there exist $x, y \in \mathbb{Z}$ such that $xa + yp = 1$. Multiply by b to get $xab + ypb = b$. Substitute kp for ab to get $(xk + yb)p = b$. Thus $p \mid b$.
($\Leftarrow$) Given that $p \mid ab \Rightarrow p \mid a$ or $p \mid b$, suppose that $a \mid p$. Then there exists $k \in \mathbb{N}$ such that $ak = p$. So $p \mid ak$, so $p \mid a$ or $p \mid k$. If $p \mid a$, then $pl = a$ for some $l \in \mathbb{N}$, in which case $alk = a$ and $lk = 1$, which implies that $k = 1$ so $a = p$. If $p \mid k$, then $k = pm$ for some $m \in \mathbb{N}$, and $apm = p$, so $am = 1$ which implies that $a = 1$. □

Remark 12 (Euclid's Statement)**.** A composite number is measured by some prime.

Euclid's Proof. Infinite regression, similar to its use in the Euclidean algorithm. □

Proposition 13. *Let n be a composite number. Then there exists a prime p such that $p \mid n$.*

Modern Proof. Since n is composite, there exist $a, b \in \mathbb{N}$ such that $1 < a, b < n$ and $n = ab$. By induction, there exists a prime p such that $p \mid b$. Thus $p \mid n$. □

Remark 14 (Euclid's Statement)**.** If a number be the least that is measured by prime numbers, it will not be measured by any other prime number except those originally measuring it.

Proposition 15 (Fundamental Theorem of Arithmetic)**.** *Let $n \in \mathbb{Z}$, $n \geq 2$. Then there exist unique prime numbers $p_1 < \cdots < p_r$ and positive integers $a_1, \ldots, a_r$ such that*

$$n = \prod_{i=1}^{r} p_i^{a_i}.$$

Proof. Let

$$X = \{m \in \mathbb{Z} \mid m \geq 2 \text{ and } m \mid n\}$$

Let $p = \min(X)$. Clearly, p is prime. If $n = p$, we are done. Otherwise, $n = pk$ for some $k \in \mathbb{Z}$. By strong induction, there exist $q_1 < \cdots < q_s$ and $b_1, \ldots, b_s$ such that $k = \prod_{i=1}^{s} q_i^{b_i}$. If $p = q_1$, set $p_i = q_i$, $a_1 = b_1 + 1$, and $a_i = b_i$ for $i > 1$, and $r = s$; otherwise set $p_1 = p$, $p_{i+1} = q_i$, $a_1 = 1$, and $a_{i+1} = b_i$, and $r = s + 1$. Now $n = u \prod_{i=1}^{r} p_i^{a_i}$. □

5. Infinitude of Primes

Remark 16. Let A be a set. We say that A *infinite* if there exists an injective function $\mathbb{N} \to A$. We say that A is *finite* if there exists a surjective function $\{1, \ldots, n\} \to A$, for some $n \in \mathbb{N}$.

Remark 17 (Euclid's Statement)**.** The prime numbers are more than any assigned multitude of prime numbers.

Proposition 18. *Let $P = \{n \in \mathbb{Z} \mid n \text{ is prime}\}$. Then P is infinite.*

Proof. Suppose that P is finite; then $P = \{p_1, \ldots, p_n\}$ for some primes p_i. Set

$$n = 1 + \prod_{i=1}^{n} p_i.$$

Since $n > p_i$ for all i, n cannot be prime; thus n is composite. Therefore there exists $p \in P$ such that $p \mid n$. This implies that $p \mid 1$, a contradiction. □

6. Problems

Problem 1. Show that the relation $\mid$ is a partial order on the set of positive integers.

Problem 2. Let $m, n \in \mathbb{Z}$ and suppose that there exist integers $x, y \in \mathbb{Z}$ such that $xm + yn = 1$. Show that $\gcd(m, n) = 1$.

Problem 3. Let $m, n \in \mathbb{N}$ and suppose that $m \mid n$. Show that $\gcd(m, n) = m$.

CHAPTER V

Archimedes on Circles and Spheres

Abstract. Disclaimer: some sections of this document were lifted from the internet, but I no longer remember which ones.

1. Precursors of Archimedes

1.1. Pythagorean Irrational Numbers. The Pythagoreans (ca. 500 B.C.) proved the existence of irrational numbers in the form of "incommensurable quantities". This tore at the fabric of their world view, based on the supremacy of whole numbers, and it is legend that the demonstrator of irrational numbers was thrown overboard at sea.

1.2. Zeno's Paradoxes. Zeno (ca. 450 B.C.) developed his famous "paradoxes of motion".

1.2.1. *The Dichotomy.* The first paradox asserts the non-existence of motion on the grounds that which is in locomotion must arrive at the half-way stage before it arrives at the goal.

1.2.2. *Achilles and the Tortoise.* The second paradox asserts that it is impossible for Achilles to overtake the tortoise when pursuing it, for he must first reach a point where the tortoise had been, but the tortoise had in the meantime moved forward.

1.2.3. *The Arrow.* The third paradox is that the flying arrow is at rest, which result follows from the assumption that time is composed of moments.

1.2.4. *The Stadium.* The fourth paradox concerns bodies which move alongside bodies in the stadium from opposite directions, from which it follows, according to Zeno, that half the time is equal to its double.

1.3. Eudoxus Method of Exhaustion. Eudoxus (ca. 370 B.C.) is remembered for two major mathematical contributions: the *Theory of Proportion*, which filled the gaps in the Pythagorean theories created by the existence of incommensurable quantities, and the *Method of Exhaustion*, which dealt with Zeno's Paradoxes. This method is based on the proposition: *If from any magnitude there be subtracted a part not less than its half, from the remainder another part not less than its half, and so on, there will at length remain a magnitude less than any preassigned magnitude of the same kind.*

Archimedes credits Eudoxus with applying this method to find that the volume of "any cone is on third part of the cylinder which has the same base with the cone and equal height."

1.4. Euclid's Elements. Euclid of Alexandria (ca. 300 B.C.) wrote *The Elements*, which may be the second most published book in history (after the Bible). The work consists of thirteen books, summarizing much of the basic

mathematics of the time, spanning plane and solid geometry, number theory, and irrational numbers.

2. Results from Euclid

Result 1. *The circumferences of two circles are to each other as their diameters.*

Using modern notation, this says that if we are given two circles with diameters D_1 and D_2, and circumferences C_1 and C_2, then

$$\frac{C_1}{C_2} = \frac{D_1}{D_2}, \quad \text{whence} \quad \frac{C_1}{D_1} = \frac{C_2}{D_2}.$$

From this, one may conclude that for any given circle, the ratio between the circumference and the diameter is a constant:

$$\frac{C}{D} = p, \quad \text{so} \quad C = pD.$$

We shall call p the *circumference constant*.

Result 2. *The areas of two circles are to each other as the squares of their diameters.*

That is, if A_1 and A_2 represent the area of circles with diameters D_1 and D_2, then

$$\frac{A_1}{A_2} = \frac{D_1^2}{D_2^2}, \quad \text{whence} \quad \frac{A_1}{D_1^2} = \frac{A_2}{D_2^2},$$

which says that there is an *area constant* for any circle:

$$\frac{A}{D^2} = k, \quad \text{so} \quad A = kD^2.$$

However, Euclid doesn't mention, and possibly doesn't realize, that p and k are related.

Result 3. *The volumes of two spheres are to each other as the cubes of their diameters.*

Thus if V_1 and V_2 are the volumes of spheres of diameter D_1 and D_2, then

$$\frac{V_1}{V_2} = \frac{D_1^3}{D_2^3}, \quad \text{whence} \quad \frac{V_1}{D_1^3} = \frac{V_2}{D_2^3};$$

again, one sees that, for again given sphere, there is a *volume constant* m such that

$$\frac{V}{D^3} = m, \quad \text{so} \quad V = mD^3.$$

Note that in each of these three cases (circumference, area, volume), the original statements by Euclid compare like units (e.g. length is to length as area is to area), whereas the modern tendency is to compare aspects of the same object (e.g. area is to length squared).

3. Measurement of a Circle

Proposition 4. *The area of any circle is equal to a right-angled triangle in which one of the sides about the right angle is equal to the radius, and the other to the circumference, of the circle.*

Let be C be the circumference, r the radius, and A the area of the circle. Let T be the area of a right triangle with legs of length r and C. Then $T = \frac{1}{2}rC$. Archimedes claims that $A = T$, so $A = \frac{1}{2}rC$.

Lemma 5. *Let h be the apothem and let Q be the perimeter of a regular polygon. Then the area of the polygon is*

$$P = \frac{1}{2}hQ.$$

Proof. Suppose the polygon has n sides, each of length b. Clearly $Q = nb$. Then the area is subdivided into n triangles of base b and height h, so

$$P = n(\frac{1}{2}hb) = \frac{1}{2}hQ.$$

□

Lemma 6. *Consider a circle of area A and let $\epsilon > 0$. Then there exists an inscribed polygon with area P_1 and a circumscribed polygon with area P_2 such that*

$$A - \epsilon < P_1 < A < P_2 < A + \epsilon.$$

Proof. Archimedes simply says: "Inscribe a square, then bisect the arcs, then bisect (if necessary) the halves and so on, until the sides of the inscribed polygon whose angular points are the points of the division subtend segments whose sum is less than the excess of the area of the circle over the triangle." □

Proof of Proposition. By double reductio ad absurdum.

Suppose that $A > T$. Then $A - T > 0$, so there exists an inscribed regular polygon with area P such that $A - P < A - T$. Thus $P > T$. If Q is the perimeter and h the apothem of the polygon, we have

$$P = \frac{1}{2}hQ < \frac{1}{2}rC = T,$$

a contradiction.

On the other hand, suppose that $A < T$. Then $T - A > 0$, so there exists a circumscribed polygon with area P such that $P - A < T - A$. Thus $P < T$. However, if Q is the perimeter and h the apothem of the polygon, we have

$$P = \frac{1}{2}hQ > \frac{1}{2}rC = T,$$

a contradiction.

Therefore, as Archimedes writes, "since then the area of the circle is neither greater nor less than [the area of the triangle], it is equal to it." □

Proposition 7. *The ratio of the circumference of any circle to its diameter is less the* $3\frac{1}{7}$ *but greater than* $3\frac{10}{71}$.

Proof. Inscribe a hexagon. Compute the area:

$$\pi = \frac{C}{D} > \frac{Q}{D} = \frac{6r}{2r} = 3.$$

Archimedes next doubles the number of vertices to obtain a regular dodecagon. The computation of its area requires accurate extraction of $\sqrt{3}$, which Archimedes estimates as

$$\left(1.732026 \approx \right)\frac{265}{153} < \sqrt{3} < \frac{1351}{780}\left(\approx 1.732051\right),$$

which is impressively close. The Archimedes continues with 24, 48, and finally 96 sides, at each stage extracting more sophisticated square roots.

Next circumscribe a hexagon and continue to 96 sides. □

In decimal notation, my calculator says that

$$3\frac{10}{71} = \frac{223}{71} \approx 3.14085 \quad < \quad \pi \approx 3.14159 \quad < \quad 3\frac{1}{7} = \frac{22}{7} \approx 3.14286.$$

4. On the Sphere and the Cylinder

The two volume work entitled *On the Sphere and the Cylinder* is Archimedes undisputed masterpiece, probably regarded by Archimedes himself as the apex of his career. These two volumes are constructed in a manner similar to Euclid's Elements, in that it proceeds from basic definitions and assumptions, through simpler known results, onto the new discoveries of Archimedes.

Among the results in this work are the following. This first describes the surface area of a sphere in terms of the area of a circle, thus comparing area to area.

Proposition 8. *The surface of any sphere is equal to four times the greatest circle in it.*

Technique of Proof. Double reductio ad absurdum: assumption that the area is more leads to a contradiction, as does assumption that the area is less. One needs to understand the area of a cone to accomplish these estimates (why?). □

Let us translate this into modern notation. Let r be the radius of the sphere and let S be its surface area. Then the radius of the greatest circle in it is πr^2. Thus Archimedes shows that

$$S = 4\pi r^2.$$

The next proposition describes the volume of a sphere in terms of the volume of a cone.

Proposition 9. *Any sphere is equal to four times the cone which has its base equal to the greatest circle in the sphere and its height equal to the radius of the sphere.*

Note that again, Archimedes has expressed the volume of the sphere in terms of the volume of a known solid; this is because the Greeks did not have modern algebraic notation. Using modern notation, we let r be the radius and let V be the volume of the sphere. The volume of the cone of radius r and height r, as determined by Eudoxus, is $\frac{1}{3}\pi r^3$. Thus

$$V = \frac{4}{3}\pi r^3.$$

In this way, Archimedes found the relationship between the circumference constant p, the area constant k (in *Measurement of a Circle*), and the volume constant m: We have

$$C = pD, \quad A = kD^2, \quad , \text{ and } V = mD^3,$$

and Archimedes has shown (in modern notation) that

$$C = \pi D \qquad \text{(that is, } p = \pi\text{)}$$

$$A = \pi r^2 = \pi\left(\frac{D}{2}\right)^2 = \frac{\pi}{4}D \qquad \text{(so } k = \frac{\pi}{4}\text{)}$$

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi\left(\frac{D}{2}\right)^3 = \frac{\pi}{6}\pi D^3 \qquad \text{(so } m = \frac{\pi}{6}\text{)}$$

From here, Archimedes now describes an astounding discovery.

Suppose we have a sphere of radius r, surface area S, and volume V. Inscribe this sphere in a right circular cylinder, whose radius would also be r and whose height would be $2r$. Then the surface area A_{cyl} of the cylinder is simply the areas of the base and top circle, plus the area of the rectangle which forms the tube of the cylinder:

$$A_{\text{cyl}} = 2(\pi r^2) + (2\pi r)(2r) = 6\pi r^2.$$

Thus

$$A_{\text{cyl}} : A_{\text{sph}} = (6\pi r^2) : (4\pi r^2) = 3 : 2.$$

Moreover, the volume of the cylinder is the area of the circular base times the height:

$$V_{\text{cyl}} = (\pi r^2)(2r) = 2\pi r^3.$$

Again, we have

$$V_{\text{cyl}} : V_{\text{sph}} = (2\pi r^3) : (\frac{4}{3}\pi r^3) = 3 : 2.$$

This so intrigued Archimedes that he requested that his tombstone be engraved with a sphere inscribed in a cylinder, together with the ratio $3 : 2$. Apparently, Marcellus, the conqueror of Syracuse, was so impressed with Archimedes, that he granted this wish.

CHAPTER VI

Diophantine Equations

1. Pythagorean Triples

A *Pythagorean triple* (a, b, c) consists of three integers $a, b, c \in \mathbb{Z}$ with $a, b \geq 1$ such that $a^2 + b^2 = c^2$.

The Babylonians produced tablets containing tables of Pythagorean triples. It is conjectured that they may have known of the formula to generate such triples: let u and v be any positive integers, and set

(a) $a = u^2 - v^2$;
(b) $b = 2uv$;
(c) $c = u^2 + v^2$.

Then

$$\begin{aligned} a^2 + b^2 &= (u^2 - v^2)^2 + (2uv)^2 \\ &= u^4 - 2u^2v^2 + v^4 + 4u^2v^2 \\ &= u^4 + 2u^2v^2 + v^4 \\ &= (u^2 + v^2)^2 \\ &= c^2. \end{aligned}$$

Thus we have:

Proposition 1. *Let $u, v \in \mathbb{Z}$ and set $a = u^2 - v^2$, $b = 2uv$, and $c = u^2 + v^2$. Then (a, b, c) is a Pythagorean triple.*

The equivalent of this scheme for generating Pythagorean triples can be found in Euclid's *Elements*, Book X, Lemma following Proposition 28. We ask if the converse is true; that is, does this method generate *all* Pythagorean triples?

2. Diophantine Equations

A *Diophantine equation* is an equation of the form

$$F(x_1, \ldots, x_n) = 0,$$

where $F(x_1, \ldots, x_n)$ is a polynomial in n variables with integer coefficients. A *solution* to a Diophantine equation is a point $(a_1, \ldots, a_n) \in \mathbb{C}^n$, where $a_i \in \mathbb{Z}$ and $F(a_1, \ldots, a_n) = 0$.

This is the modern definition. However, Diophantus looked for rational solutions to polynomial equations with integer (or rational) coefficients. We note that a rational solution to a polynomial equation produces an integer solution to a modified polynomial equation, obtained by clearing the denominators; that is, multiply the expression $F(a_1, \ldots, a_n) = 0$ by the least common multiple of the highest powers of the denominators of $a_1, \ldots, a_n$ to appear in the expression.

Example 2. Pythagorean triples are integer solutions to the polynomial equation $x^2 + y^2 = z^2$. Suppose (a, b, c) is such a solution, with $a, b, c \in \mathbb{Z}$. Then $(\frac{a}{c}, \frac{b}{c})$ is a rational solution to the equation $x^2 + y^2 = 1$. Thus the problem of finding integer solutions to $x^2 + y^2 = z^2$ is equivalent to the problem of finding rational solutions to $x^2 + y^2 = 1$.

Example 3. Fermat's last theorem essentially states that there are no nontrivial integer solutions to the equation $x^n + y^n = z^n$ for $n \geq 3$. Again, this is equivalent to the nonexistence of nontrivial rational solutions to $x^n + y^n = 1$ for $n \geq 3$.

Example 4. Fix $m, n \in \mathbb{Z}$, and let $d = \gcd(m, n)$. The Euclidean algorithm produces unique solutions to the Diophantine equation $mx + ny = d$.

An *plane algebraic curve* is the subset of $\mathbb{C}^2$ which is the set of points $(a, b) \in \mathbb{C}^2$ such that $F(a, b) = 0$ for some polynomial $F(x, y)$ with coefficients in $\mathbb{C}$. We say that the curve is *defined over* $\mathbb{Q}$ if these coefficients are in $\mathbb{Q}$. The *degree* of the curve is the degree of the polynomial F. A *rational point* on the curve is a point on the curve with rational coordinates.

Diophantus studied plane algebraic curves, and looked for rational solutions to such polynomial equations, which is to say, he attempted to find rational points on the associated algebraic curve. Keep in mind that the notation used by Diophantus was very dissimilar to that used today.

Example 5. We see that (a, b, c) is a Pythagorean triple if and only if $(\frac{a}{c}, \frac{b}{c})$ is a rational point on the curve $x^2 + y^2 = 1$.

Example 6. Fermat's Last Theorem amounts to the claim that $(1, 0)$ and $(0, 1)$ are the only rational points on the curve $x^n + y^n = 1$.

3. Generation of Pythagorean Triples

One technique used by Diophantus to find rational points on a curve was to find an apparent solution P and intersect the curve with lines through P which have rational slope. That this works for conic sections is exemplified by the following propositions.

Proposition 7. *Let $f(x) = ax^2 + bx + c$, where $a, b, c \in \mathbb{Q}$. If x_1, x_2 satisfy $f(x) = 0$, and $x_1 \in \mathbb{Q}$, then $x_2 \in \mathbb{Q}$.*

Proof. Suppose $x_2 \neq x_1$. Set $d = \sqrt{b^2 - 4ac}$. Then

$$x_1 = \frac{-b + ud}{2a} \quad \text{and } x_2 = \frac{-b - ud}{2a},$$

where $u = 1$ or $u = -1$. Therefore $d = u(2ax_1 + b) \in \mathbb{Q}$. Therefore $x_2 \in \mathbb{Q}$. □

Proposition 8. *Let $P = (-1, 0)$ and $Q = (a, b)$ with $a^2 + b^2 = 1$ and $a > -1$. Then P and Q are distinct points on the unit circle $x^2 + y^2 = 1$, and Q is a rational point if and only if the slope of line through P and Q is rational.*

Proof. Let m be the slope of the line through P and Q; then

$$m = \frac{b}{a + 1},$$

and the equation of the line through P and Q is $y = m(x + 1)$.

If Q is rational, this means that $a, b \in \mathbb{Q}$, so $m = \frac{b}{a+1} \in \mathbb{Q}$.

On the other hand, suppose that the slope is rational. The x-coordinate of the intersection of the curve and the line satisfies

$$x^2 + (m(x + 1))^2 = 1.$$

This is a quadratic equation whose solutions, for our given m, are $x = -1$ and $x = a$; therefore, a is rational by Proposition 7. □

The problem of finding Pythagoreans triples is equivalent to the problem of finding rational points on the curve $x^2 + y^2 = 1$. Diophantus realized that all such points could be obtained by running a line with rational slope through the point $P = (-1, 0)$ and taking the point of intersection with the unit circle. We compute these points as follows.

Let $m \in \mathbb{Q}$; the line with slope m through P is $y = m(x + 1)$. Let Q be the other point of intersection of this line with the unit circle. Substituting $m(x+1)$ for y in the equation of the unit circle gives $x^2 + (m(x + 1))^2 = 1$, or

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

This quadratic equation has solutions

$$x = \frac{-2m^2 \pm \sqrt{4m^4 - 4(m^4 - 1)}}{2(m^2 + 1)} = \frac{-m^2 \pm 1}{m^2 + 1},$$

so the solution that produces P is $x = -1$, and the solution that produces Q is

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

Substitute this into the line to get

$$\boxed{y = \frac{2m}{1+m^2}.}$$

We have shown:

Proposition 9. *Let* $\mathbb{U} = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ *and let* $P = (-1, 0)$. *The function*

$$\phi : \mathbb{Q} \to \mathbb{U} \quad \text{given by } \phi(m) = \left(\frac{1-m^2}{1+m^2}, \frac{2m}{1+m^2}\right)$$

produces a bijective correspondence between the rational numbers and the rational points (other than P*) on the unit circle.*

Now plug these values for x and y into the equation of the circle and get

$$\left(\frac{1-m^2}{1+m^2}\right)^2 + \left(\frac{2m}{1+m^2}\right)^2 = 1,$$

therefore

$$(1-m^2)^2 + 4m^2 = (1+m^2)^2.$$

Let $m \in \mathbb{Q}$ be positive; then there exist positive $u, v \in \mathbb{Z}$ such that $m = \frac{v}{u}$. Then, substituting this into the above formula and clearing the denominators by multiplying by u^4, we obtain

$$(u^2 - v^2)^2 + (2uv)^2 = (u^2 + v^2)^2.$$

This shows:

Theorem 10 (Diophantus' Theorem)**.** *Let* (a, b, c) *be a Pythagorean triple. Then there exist* $u, v \in \mathbb{Z}$ *such that* $a = u^2 - v^2$, $b = 2uv$, *and (consequently)* $c = u^2 + v^2$.

4. Cubic Equations

Diophantus also applied this technique to cubic equations in two variables, using the fact that the generic degree three polynomial in one variable has three solutions, and if two of them are rational, then so is the third.

Given a degree three curve defined over $\mathbb{Q}$ by the equation $F(x, y) = 0$, the intersection of the curve with a line $y = mx+b$ gives an equation $F(x, mx+b) = 0$. If two rational solutions are known, then the third solution must also be rational.

Suppose we find one rational point $P = (a, b)$ on the curve. If we select a nearby point on the curve and let it approach P, the secant line between the points approaches the tangent line $y = mx + b$. Then m is rational, and if this tangent line intersects the curve in another point, the other point will also be rational. This is because a is a double root of $F(x, mx + b) = 0$.

As an aside, we note that this technique re-emerged in the early 19^{th} century in the following context. In attempting to compute the arclength along an ellipse, Niels Henrik Abel discovered certain integrals, known as *elliptic integrals*, with the property that the natural domain of the inverse of the antiderivative was a torus as opposed to the Riemann sphere (this is the traditional name for the complex plane together with a point at ∞). This developed into the study of *elliptic curves*, which are curves defined by an equation of the form $y^2 = f(x)$, where $f(x)$ is a cubic polynomial.

In 1835, Carl Gustav Jacob Jacobi created a type of addition of the points on an elliptic curve, called the *chord-tangent law*, which can be defined in terms of taking lines through points and intersecting them with the curve. Under this addition, the sum of rational points is also rational, so the set of rational points form an algebraic system known as an *abelian group*.

Example 11. Find three rational points on the curve $y^2 = x^3 - 3x^2 + 3x + 1$.

Solution. We see that $(0, 1)$ and $(0, -1)$ are solutions. Let $P = (0, 1)$; we would like to find the line tangent to the curve through the point P. Using implicit differentiation (which was not available to Diophantus), we compute that

$$2y\frac{dy}{dx} = 3x^2 - 6x + 3,$$

so

$$\frac{dy}{dx} = \frac{3(x^2 - 2x + 1)}{2y}.$$

Set

$$m = \frac{dy}{dx}\,|_P = \frac{3}{2}.$$

If $P = (x_0, y_0) = (0, 1)$, the tangent line is

$$y = m(x - x_0) + y_0 = \frac{3}{2}x + 1.$$

Substitute this into the equation of the curve to get

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

Solving for x gives $x = \frac{21}{4}$. Applying this to the line produces $y = \frac{71}{8}$. This is rational; thus $(\frac{21}{4}, \frac{71}{8})$ is a rational point on the curve. □

5. Problems

Problem 1. The equation $y^2 = x^3 - ax + b$ defines an *elliptic curve*.

- **(a)** Use calculus to find all points on the curve with horizontal or vertical tangents.
- **(b)** Let $a = 12$ and $b = 25$. Take a horizontal tangent and intersect it with this curve to find another rational point.
- **(c)** Let $a = 2$ and $b = 0$. Find three rational points on this curve.

CHAPTER VII

Modular Arithmetic

Abstract. Congruence relations were formalized by Gauss at the beginning of the nineteenth century; however, important components of the theory were realized by the ancient Greeks, Arabs, and Chinese. We investigate this, with an eye towards understanding the Chinese Remainder Theorem.

1. Review of Integer Properties

Fact 1. Division Algorithm for Integers
Let $m, n \in \mathbb{Z}$. There exist unique integers $q, r \in \mathbb{Z}$ such that

$$n = qm + r \qquad \text{and} \qquad 0 \leq r < m.$$

Definition 2. Let $m, n \in \mathbb{Z}$. We say that m *divides* n, and write $m \mid n$, if there exists an integer k such that $n = km$.

Definition 3. Let $m, n \in \mathbb{Z}$. A *greatest common divisor* of m and n, denoted $\gcd(m, n)$, is a positive integer d such that

(1) $d \mid m$ and $d \mid n$;
(2) If $e \mid m$ and $e \mid n$, then $e \mid d$.

Fact 4. Euclidean Algorithm for Integers
Let $m, n \in \mathbb{Z}$. Then there exists a unique $d \in \mathbb{Z}$ such that $d = \gcd(m, n)$, and there exist integers $x, y \in \mathbb{Z}$ such that

$$xm + yn = d.$$

Definition 5. An integer $p \geq 2$, is called *prime* if

$$a \mid p \Rightarrow a = 1 \text{ or } a = p, \qquad \text{where } a \in \mathbb{N}.$$

An integer $n \geq 2$ is called *composite* if it is not prime.

Fact 6. Fundamental Theorem of Arithmetic
Let $n \in \mathbb{Z}$, $n \geq 2$. Then there exist unique prime numbers $p_1 < \cdots < p_r$ and positive integers $a_1, \ldots, a_r$ such that

$$n = \prod_{i=1}^{r} p_i^{a_i}.$$

2. Congruence Modulo n

Proposition 7. *Let $n \in \mathbb{Z}$ with $n \geq 2$, and let $a, b, c \in \mathbb{Z}$. Then*

- **(a)** *$a \equiv a \pmod{n}$ (Reflexivity);*
- **(b)** *if $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$ (Symmetry);*
- **(c)** *if $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$ (Transitivity).*

Proof.
(*Reflexivity*) Note that $0 \cdot n = 0 = a - a$; thus $n \mid (a - a)$, so $a \equiv a$. Therefore $\equiv$ is reflexive.
(*Symmetry*) Let $a, b \in \mathbb{Z}$. Suppose that $a \equiv b$; then $n \mid (a - b)$. Then there exists $k \in \mathbb{Z}$ such that $nk = a - b$. Then $n(-k) = b - a$, so $n \mid (b - a)$. Thus $b \equiv a$. Similarly, $b \equiv a \Rightarrow a \equiv b$. Therefore $\equiv$ is symmetric.
(*Transitivity*) Let $a, b, c \in \mathbb{Z}$, and suppose that $a \equiv b$ and $b \equiv c$. Then $nk = a - b$ and $nl = b - c$ for some $k, l \in \mathbb{Z}$. Then $a - c = nk - nl = n(k - l)$, so $n \mid (a - c)$. Thus $a \equiv c$. Therefore $\equiv$ is transitive. □

Proposition 8. *Let $n \in \mathbb{Z}$ with $n \geq 2$. Let $a, b \in \mathbb{Z}$. Then $a \equiv b \pmod{n}$ if and only if a and b have the same remainder upon division by n.*

Proof. By the division algorithm, there exist unique integer q_1, q_2, r_1, r_2 such that

$$a = nq_1 + r_1 \quad \text{with } 0 \leq r_1 \leq n$$

and

$$b = nq_2 + r_2 \quad \text{with } 0 \leq r_2 \leq n.$$

Thus $a - b = n(q_1 - q_1) + (r_1 - r_2)$.

If $a \equiv b \pmod{n}$, then $n \mid (a - b)$, so $a - b = kn$ for some $k \in \mathbb{Z}$. Thus $kn = n(q_1 - q_2) + (r_1 - r_2)$, so $r_1 - r_2 = n(k - q_1 + q_2)$; that is, $r_1 - r_2$ is a multiple of n. But subtracting the inequalities bounding the remainders shows that $-n < r_1 - r_2 < n$, and the only multiple of n in this range is zero. So $r_1 - r_2 = 0$, whence $r_1 = r_2$.

On the other hand, if $r_1 = r_2$, then we have $a - b = n(q_1 - q_2)$, so $a - b$ is divisible by n, and $a \equiv b \pmod{n}$. □

Proposition 9. *Let $n \in \mathbb{Z}$ with $n \geq 2$. Let $a, b, c, d \in \mathbb{Z}$ with $a \equiv c$ and $b \equiv d$. Then*

- **(a)** *$a + b \equiv c + d \pmod{n}$;*
- **(b)** *$ab \equiv cd \pmod{n}$.*

Proof. All equivalences will be taken modulo n. Since $a \equiv c$ and $b \equiv d$, there exist $p, q \in \mathbb{Z}$ such that $a - c = pn$ and $b - d = qn$.

Now $a+b = c+pn+d+qn = (c+d)+n(p+q)$, so $(a+b)-(c+d) = n(p+q)$, whence $a + b \equiv c + d$.

Similarly, $ab = (c+pn)(d+qn) = cd+cqn+dpn+pqn^2 = cd+n(cq+dp+pqn)$, whence $ab-cd = n(cq+dp+pqn)$, so $ab-cd$ is divisible by n. Thus $ab \equiv cd$. □

3. Casting Out n's

The process of *casting out n's* involves subtracting n from a number until one arrives at a number less than n. Clearly, this number is the remainder upon division by n, so it is related to modular arithmetic.

The method of casting out n's, together with decimal notation, led Arabs of 1500 years ago to discover certain divisibility criteria. We demonstrate this in modern notation.

Fix $n \in \mathbb{Z}$ with $n \geq 0$. For $a \in \mathbb{Z}$, let $\overline{a}$ denote the remainder when a is divide by n. The last proposition states that $\overline{a+b} \equiv \overline{a} + \overline{b}$ and $\overline{ab} \equiv \overline{a}\overline{b}$, modulo n.

If $d_0, d_1, \ldots, d_r$ are the digits of $a \in \mathbb{N}$, then

$$a = \sum_{i=0}^{r} d_i 10^i.$$

The idea of casting out n's revolves around the fact that

$$a \equiv \sum_{i=0}^{r} \overline{d_i 10}^i \pmod{n}.$$

Proposition 10. Casting Out 3's and 9's
Let $a \in \mathbb{Z}$ be a positive integer with decimal expansion

$$a = \sum_{i=0}^{k} d_i 10^i,$$

where $0 \leq d_i \leq 9$ for $i = 0, \ldots, k$. Set

$$s = \sum_{i=0}^{k} d_i$$

Let $n = 3$ or $n = 9$. Then a is divisible by n if and only if s is divisible by n.

Proof. Let $n = 3$ or $n = 9$ and consider equivalence modulo n. Note that $10 \equiv 1 \pmod{n}$ for $n = 3$ or $n = 9$. Then we have

$$\begin{aligned} a &= \overline{\sum_{i=0}^{k} d_i 10^i} \\ &\equiv \sum_{i=0}^{k} d_i \overline{10}^i \\ &\equiv \sum_{i=0}^{k} d_i \quad \text{because} \quad \overline{10} = 1 \\ &= s. \end{aligned}$$

So a and s have the same remainder upon division by n, and in particular a is divisible by n if and only if s is divisible by n. □

Proposition 11. Casting Out 11's
Let $a \in \mathbb{Z}$ be a positive integer with decimal expansion

$$a = \sum_{i=0}^{k} d_i 10^i,$$

where $0 \leq d_i \leq 9$ for $i = 0, \ldots, k$. Set

$$s = \sum_{i=0}^{k} (-1)^i d_i$$

Let $n = 11$. Then a is divisible by n if and only if s is divisible by n.

Proof. Let $n = 11$. In this case, $10 \equiv -1 \pmod{n}$. We have

$$\begin{aligned} a &= \sum_{i=0}^{k} d_i 10^i \\ &\equiv \sum_{i=0}^{k} d_i \overline{10}^i \\ &\equiv \sum_{i=0}^{k} d_i \overline{-1}^i \\ &\equiv \sum_{i=0}^{k} (-1)^i d_i \\ &= s. \end{aligned}$$

Thus a is divisible by n if and only if s is divisible by n. □

4. Chinese Remainder Theorem

Proposition 12. *Let $a, b, m, n \in \mathbb{Z}$ such that* $\gcd(m, n) = 1$.
Then there exists $c \in \mathbb{Z}$ such that

- $c \equiv a \mod m$;
- $c \equiv b \mod n$.

Proof. There exist $x, y \in \mathbb{Z}$ such that $mx + ny = 1$. Let $c = mxb + nya$. Then

$$c - a = mxb + nya - a = mxb + (ny - 1)a = mxb - mxa,$$

so m divides $c - a$; thus $c \equiv a \mod m$. Also

$$c - b = mxb + nya - b = (mx - 1)b + nya = -nyb + nya,$$

so n divides $c - b$; thus $c \equiv b \mod n$. □

Example 13. Let $m = 104$, $n = 231$, $a = 11$, and $b = 23$. Find $c \in \mathbb{Z}$ with $0 \leq c < mn$ such that $c \equiv a \pmod m$ and $c \equiv b \pmod n$.

Solution. First we use the Euclidean algorithm to write $mx + yn = d$. We have

$$\begin{aligned} 231 &= 104 \cdot 2 + 23 \\ 104 &= 23 \cdot 4 + 12 \\ 23 &= 12 \cdot 1 + 11 \\ 12 &= 11 \cdot 1 + 1 \\ 11 &= 1 * 11 + 0 \end{aligned}$$

Thus

$$\begin{aligned} 1 &= (-1)11 + 12 \\ &= (2)12 + (-1)23 \\ &= (-9)23 + (2)104 \\ &= (20)104 + (-9)231 \end{aligned}$$

That is, $x = 20$, $y = -9$, and $d = 1$,

Now set

$$c = mxb + nya \pmod{24024} = 24971 \pmod{24024} = 947.$$

□

CHAPTER VIII

The Fibonacci Sequence

ABSTRACT. Sequences play an important role in modern mathematics, and one of the first to investigate them was Leonardo Fibonacci in the twelfth century A.D. We investigate the famous sequence which perpetuates his name.

1. Recursively Defined Sequences

Definition 1. Let X be a set. A *sequence* in X is a function $a : \mathbb{N} \to X$. We normally write a_n to mean $a(n)$, and the entire function is often denoted by $(a_n)_{n=1}^{\infty}$, or simply as (a_n).

Definition 2. Let (a_n) be a sequence in $\mathbb{R}$, and let $L \in \mathbb{R}$. We say that (a_n) *converges* to L, or that L is the *limit* of (a_n), if

$$\text{for every } \epsilon > 0 \text{ there exists } N \in \mathbb{N} \text{ such that } n \geq N \Rightarrow |a_n - L| < \epsilon.$$

In this case we write $\lim a_n = L$.

We assume familiarity with the standard properties, and focus on recursively defined sequences. Suppose that we set $a_0 = C$, a fixed constant value, select a function $f : \mathbb{R} \to \mathbb{R}$, and set $a_{n+1} = f(a_n)$ for every n. This uniquely defines a sequence (a_n) of real numbers.

Now it is clear that if we obtain a new sequence (a_{n+1}) from (a_n) by shifting, the limit (should it exist) does not change: $\lim a_{n+1} = \lim a_n$. If (a_n) is a recursively defined sequence such that $a_{n+1} = f(a_n)$ for some *continuous* function f, then $\lim a_{n+1} = f(\lim a_n)$, so if $L = \lim a_n$, we have $L = f(L)$. We use this fact to analyze recursively defined sequences (accept that the following sequences do converge; proving this is typically harder than computing the limit of a recursively defined sequence).

Example 3. Define a sequence (a_n) by $a_0 = 1$ and $a_{n+1} = \frac{a_n}{2}$. Find $\lim a_n$.

Solution. The first few terms of the sequence are $a_0 = 1$, $a_1 = \frac{1}{2}$, $a_2 = \frac{1/2}{2} = \frac{1}{4}$, $a_3 = \frac{1/4}{2} = \frac{1}{8}$, and so forth; we see that this sequence could have been given as $a_n = \frac{1}{2^n}$. In fact, if $L = \lim a_n$, then $L = \frac{L}{2}$, so $2L = L$, so $L = 0$. □

Example 4. Define a sequence (a_n) by $a_0 = 1$ and $a_{n+1} = \frac{a_n+1}{3}$. Find $\lim a_n$.

Solution. In this case, $a_0 = 1$, $a_1 = \frac{2}{3}$, $a_2 = \frac{5}{9}$, $a_3 = \frac{14}{27}$, $a_4 = \frac{41}{81}$, and so forth. We believe that $a_n = \frac{(3^n+1)/2}{3^n}$; the sequence certainly seems to be approaching $\frac{1}{2}$. In fact, with $L = \lim a_n$, we have $L = \frac{L+1}{3}$, so $3L = L + 1$, so $2L = 1$, and $L = \frac{1}{2}$. □

Example 5. Define a sequence (a_n) by $a_0 = 1$ and $a_{n+1} = \sqrt{1 + a_n}$. Find $\lim a_n$.

Solution. This sequence formalizes the repeated square root

$$\sqrt{1 + \sqrt{1 + \sqrt{1 + \ldots}}}.$$

We have $L = \sqrt{1 + L}$, so $L^2 = 1 + L$, and $L^2 - L - 1 = 0$. Noting the limit must be positive, the quadratic formula gives $L = \frac{1+\sqrt{5}}{2}$. That is, L is the golden ratio Φ. The sequence increases to this upper bound. □

Example 6. Define a sequence (a_n) by $a_1 = 1$ and $a_{n+1} = 1 + \frac{1}{a_n}$. Find $\lim a_n$.

Solution. This sequence formalizes the repeated fraction

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}}.$$

Let's compute the first few terms of this sequence; we will see an interesting pattern.

- $a_1 = 1$
- $a_2 = 1 + \frac{1}{1} = \frac{1+1}{1} = 2$
- $a_3 = 1 + \frac{1}{2} = \frac{2+1}{2} = \frac{3}{2}$
- $a_4 = 1 + \frac{2}{3} = \frac{3+2}{3} = \frac{5}{3}$
- $a_5 = 1 + \frac{3}{5} = \frac{5+3}{5} = \frac{8}{5}$
- $a_6 = 1 + \frac{5}{8} = \frac{8+5}{8} = \frac{13}{8}$

We see that, in each case, we add the numerator and denominator and put it over the previous numerator.

We compute that if $L = \lim a_n$, then $L = 1 + \frac{1}{L}$, so $L^2 = L + 1$, so $L^2 - L - 1 = 0$, and $L = \frac{1+\sqrt{5}}{2}$. Actually, the sequence jumps back and forth around Φ, with the even terms less than Φ and the odd terms greater than Φ. □

2. Fibonacci Sequence

Definition 7. Define a sequence (F_n) by setting $F_1 = 1$, $F_2 = 1$, and

$$F_{n+2} = F_n + F_{n+1}.$$

Then (F_n) is known as the *Fibonacci sequence*, after the 12^{th} century mathematician Fibonacci, who discovered the sequence while investigating the breeding of rabbits.

The first few terms of the Fibonacci sequence are

$$1, 1, 2, 3, 5, 8, 13, 21, 44, 65, 109, 174, 283, 475, \ldots$$

Define a sequence (a_n) by $a_0 = 1$ and $a_n = \frac{F_{n+1}}{F_n}$. Then $a_1 = 1$, $a_2 = 2$, $a_3 = \frac{3}{2}$, $a_4 = \frac{5}{3}$; look familiar? Now

$$a_{n+1} = \frac{F_{n+2}}{F_{n+1}} = \frac{F_{n+1} + F_n}{F_{n+1}} = 1 + \frac{1}{a_n};$$

so as we have already seen,

$$\lim \frac{F_{n+1}}{F_n} = \frac{1 + \sqrt{5}}{2}.$$

The golden ratio is also involved in the following *generating function* for the Fibonacci sequence:

Proposition 8.

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Solution. The *golden ratio* is the positive solution to the equation $x^2 - x - 1 = 0$; the quadratic formula gives the roots as $\frac{1\pm\sqrt{5}}{2}$. Set

$$\Phi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \Psi = \frac{1-\sqrt{5}}{2}.$$

then Φ and Ψ satisfy the above equation, which produces these identities:

- $\Phi + 1 = \Phi^2$;
- $\Phi - 1 = \frac{1}{\Phi}$;
- $\Psi + 1 = \Phi^2$;
- $\Psi - 1 = \frac{1}{\Psi}$;
- $\Psi = -\frac{1}{\Phi} = 1 - \Phi$;
- $\Phi - \Psi = \sqrt{5}$.

In light of this, what we wish to show can be rewritten as

$$F_n = \frac{1}{\sqrt{5}}\left(\Phi^n - \Psi^n\right).$$

We have $F_1 = 1$ and plugging 1 into the above expression produces

$$\frac{1}{\sqrt{5}}\left(\Phi - \Psi\right) = \frac{\sqrt{5}}{\sqrt{5}} = 1;$$

therefore the formula is true for $n = 1$.

By strong induction, assume that for $n \geq 3$ we have

$$F_{n-2} = \frac{1}{\sqrt{5}}\left(\Phi^{n-2} - \Psi^{n-2}\right);$$
$$F_{n-1} = \frac{1}{\sqrt{5}}\left(\Phi^{n-1} - \Psi^{n-1}\right),$$

Then

$$\begin{aligned}
F_n &= F_{n-2} + F_{n-1} \\
&= \frac{1}{\sqrt{5}}\left(\Phi^{n-2} - \Psi^{n-2}\right) + \frac{1}{\sqrt{5}}\left(\Phi^{n-1} - \Psi^{n-1}\right) \\
&= \frac{1}{\sqrt{5}}\left((\Phi^{n-2} + \Phi^{n-1}) - (\Psi^{n-2} + \Psi^{n-1})\right) \\
&= \frac{1}{\sqrt{5}}\left(\Phi^{n-2}(1+\Phi) - \Psi^{n-2}(1+\Psi)\right) \\
&= \frac{1}{\sqrt{5}}\left(\Phi^{n-2}(\Phi^2) - \Psi^{n-2}(\Psi^2)\right) \\
&= \frac{1}{\sqrt{5}}\left(\Phi^n - \Psi^n\right).
\end{aligned}$$

This completes the proof. □

3. Cauchy Sequences

We now supply a formal proof that the sequence of ratios of the Fibonacci numbers is a Cauchy sequence, and so it does in fact converge.

Definition 9. Let (a_n) be a sequence of real numbers. We say that (a_n) is a *Cauchy sequence* if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$m, n \geq N \quad \Rightarrow \quad |a_m - a_n| < \epsilon.$$

The proof of the next theorem may be found in books on real analysis.

Theorem 10. (Cauchy Convergence Criterion)
A sequence of real numbers converges if and only if it is a Cauchy sequence.

Proposition 11. *Let (a_n) be a sequence satisfying*

$$|a_{n+1} - a_n| < \frac{1}{2^n}$$

for all $n \in \mathbb{N}$. Then (a_n) is a Cauchy sequence.

Lemma 12. *Let $m, n \in \mathbb{N}$ with $2 < m < n$. Then*

$$\sum_{i=m+1}^{n} \frac{1}{2^i} < \frac{1}{2^m} < \frac{1}{m}.$$

Proof of Lemma. We prove the first inequality by induction on $k = n - m$. If $k = 1$, then our statement reads $\frac{1}{2^{m+1}} < \frac{1}{2^m}$, which is true.

Suppose that our proposition is true for differences of size $k - 1$. Then

$$\sum_{i=m+2}^{n} \frac{1}{2^i} < \frac{1}{2^{m+1}}.$$

Adding $\frac{1}{2^{m+1}}$ to both sides gives

$$\sum_{i=m+1}^{n} \frac{1}{2^i} < \frac{2}{2^{m+1}} = \frac{1}{2^m}.$$

For the second inequality, it suffices to show that for $m > 2$ we have $m < 2^m$. For $m = 3$, we have $3 < 4$. By induction, $m - 1 < 2^{m-1}$. Then $m < 2^{m-1} + 1 < 2^{m-1} + 2^{m-1} = 2^m$. □

Proof of Proposition. Let $\epsilon > 0$ and let $N \in \mathbb{N}$ be so large that $\frac{1}{\epsilon} < N$. Let $m, n > N$; assume that $n > m$. Then

$$\begin{aligned} |a_n - a_m| &= |a_n - a_{n-1} + a_{n-1} - a_{n-2} + \cdots + a_{m+1} - a_m| \\ &\leq |a_n - a_{n-1}| + \cdots + |a_{m+1} - a_m| \\ &< \frac{1}{2^{n-1}} + \cdots + \frac{1}{2^m} \\ &< \frac{1}{2^{m-1}} \\ &< \frac{1}{m-1} \leq \frac{1}{N} < \epsilon. \end{aligned}$$

This shows that (a_n) is a Cauchy sequence. □

Proposition 13. *Define a sequence* (a_n) *by*

$$a_n = \frac{F_{n+1}}{F_n}.$$

Then (a_n) *is a Cauchy sequence which converges to* $\frac{1+\sqrt{5}}{2}$.

Proof. To show that (a_n) is a Cauchy sequence, it suffices to show that

$$|a_{n+1} - a_n| < \frac{1}{2^{n-1}}.$$

To do this, we first show that $F_n F_{n+1} > 2^{n-1}$ for $n \geq 3$. For $n = 3$, we have $F_3 F_4 = 2 \cdot 3 > 4$. By induction, assume that $F_{n-1} F_n > 2^{n-2}$. Clearly (F_n) is a nondecreasing sequence, so

$$F_n F_{n+1} = F_n^2 + F_n F_{n-1} \geq 2 F_n F_{n-1} > 2^{n-1}.$$

Next we show that $|F_n F_{n+2} - F_{n+1}^2| = 1$ for $n \geq 1$. For $n = 1$, we have $|F_1 F_3 - F_2^2| = 2 - 1 = 1$. By induction, assume that $|F_{n-1} F_{n+1} - F_n^2| = 1$. Then

$$\begin{aligned}
|F_n F_{n+2} - F_{n+1}^2| &= |F_n(F_n + F_{n+1}) - F_{n+1}^2| \\
&= |F_n^2 + F_n F_{n+1} - F_{n+1}^2| \\
&= |F_n^2 - F_{n+1}(F_{n+1} - F_n)| \\
&= |F_n^2 - F_{n+1} F_{n-1}| \\
&= 1.
\end{aligned}$$

Now

$$\begin{aligned}
|a_{n+1} - a_n| &= \left| \frac{F_{n+2}}{F_{n+1}} - \frac{F_{n+1}}{F_n} \right| \\
&= \left| \frac{F_{n+2} F_n - F_{n+1}^2}{F_n F_{n+1}} \right| \\
&= \left| \frac{1}{F_n F_{n+1}} \right| \\
&< \frac{1}{2^{n-1}}.
\end{aligned}$$

Since (a_n) is a Cauchy sequence, it converges; let $L = \lim(a_n)$. Since a_n is positive for all n, $L \geq 0$. Now

$$a_{n+1} = \frac{F_{n+2}}{F_{n+1}} = \frac{F_n + F_{n+1}}{F_{n+1}} = 1 + \frac{F_n}{F_{n+1}} = 1 + \frac{1}{a_n}.$$

Taking the limit of both sides of this equation, we have $L = 1 + \frac{1}{L}$. Thus

$$L^2 - L - 1 = 0.$$

The positive solution to this quadratic equation is

$$L = \frac{1 + \sqrt{5}}{2}.$$

□

Proposition 14. *Let $b \in \mathbb{R}$, $b \geq 1$, and define a sequence (G_n) by $G_1 = 1$, $G_2 = 1$, and $G_{n+2} = G_n + bG_{n+1}$. Define a sequence (c_n) by*

$$c_n = \frac{G_{n+1}}{G_n}.$$

Then (c_n) is a Cauchy sequence.

Proof. To show that (c_n) is a Cauchy sequence, it suffices to show that

$$|c_{n+1} - c_n| < \frac{b}{2^{n-1}}.$$

To do this, we first show that $G_nG_{n+1} > 2^{n-1}$ for $n \geq 3$. For $n = 3$, we have $G_3G_4 = (b+1)(b^2+b+1) > 4$. By induction, assume that $G_{n-1}G_n > 2^{n-2}$. Clearly (G_n) is a nondecreasing sequence, so

$$G_nG_{n+1} = bG_n^2 + G_nG_{n-1} \geq G_n^2 + G_nG_{n-1} \geq 2G_nG_{n-1} > 2^{n-1}.$$

Next we show that $|G_nG_{n+2} - G_{n+1}^2| = b$ for $n \geq 1$. For $n = 1$, we have $|G_1G_3 - G_2^2| = b + 1 - 1 = b$. By induction, assume that $|G_{n-1}G_{n+1} - G_n^2| = b$. Then

$$\begin{aligned}
|G_nG_{n+2} - G_{n+1}^2| &= |G_n(G_n + bG_{n+1}) - G_{n+1}^2| \\
&= |G_n^2 + bG_nG_{n+1} - G_{n+1}^2| \\
&= |G_n^2 - G_{n+1}(G_{n+1} - bG_n)| \\
&= |G_n^2 - G_{n+1}G_{n-1}| \\
&= b.
\end{aligned}$$

Now

$$\begin{aligned}
|c_{n+1} - c_n| &= \left|\frac{G_{n+2}}{G_{n+1}} - \frac{G_{n+1}}{G_n}\right| \\
&= \left|\frac{G_{n+2}G_n - G_{n+1}^2}{G_nG_{n+1}}\right| \\
&= \left|\frac{b}{G_nG_{n+1}}\right| \\
&< \frac{b}{2^{n-1}}.
\end{aligned}$$

Thus (c_n) is a Cauchy sequence. □

CHAPTER IX

Cubic Equations and Quartic Equations

1. The Story

Various solutions for solving quadratic equations $ax^2 + bx + c = 0$ have been around since the time of the Babylonians. A few methods for attacking special forms of the cubic equation $ax^3 + bx^2 + cx + d = 0$ had been investigated prior to the discovery and development of a general solution to such equations, beginning in the fifteenth century and continuing into the sixteenth century, A.D. This story is filled with bizarre characters and plots twists, which is now outlined before describing the method of solution.

The biographical material here was lifted wholesale from the MacTutor History of Mathematics website, and then editted. Other material has been derived from Dunham's *Journey through Genius*.

1.1. Types of Cubics. Zero and negative numbers were not used in fifteenth century Europe. Thus, cubic equations were viewed to be in different types, depending on the degrees of the terms and there placement with respect to the equal sign.

- $x^3 + mx = n$ "cube plus cosa equals number"
- $x^3 + mx^2 = n$ "cube plus squares equals number"
- $x^3 = mx + n$ "cube equals cosa plus number"
- $x^3 = mx^2 + n$ "cube equals squares plus number"

Mathematical discoveries at this time were kept secret, to be used in public "debates" and "contests". For example, the method of depressing a cubic (eliminating the square term by a linear change of variable) was discovered independently by several people. The more difficult problem of solving the depressed cubic remained elusive.

1.2. Luca Pacioli. In 1494, Luca Pacioli published *Summa de arithmetica, geometria, proportioni et proportionalita*. The work gives a summary of the mathematics known at that time although it shows little in the way of original ideas. The work studies arithmetic, algebra, geometry and trigonometry and, despite the lack of originality, was to provide a basis for the major progress in mathematics which took place in Europe shortly after this time. The book admittedly borrows freely from Euclid, Boethius, Sacrobosco, Fibonacci, et cetera.

In this book, Pacioli states that the solution of the cubic is impossible.

1.3. Scipione del Ferro. The first known mathematician to produce a general solution to a cubic equation is Sipione del Ferro. He knew that the problem of solving the general cubic could be reduced to solving the two cases $x^3 + mx = n$ and $x^3 = mx + n$, where m and n are positive numbers, and del

Ferro may have solved both cases; we do not know for certain, because his results were never published.

We know that del Ferro was appointed as a lecturer in arithmetic and geometry at the University of Bologna in 1496 and that he retained this post for the rest of his life. No writings of del Ferro have survived. We do know however that he kept a notebook in which he recorded his most important discoveries. This notebook passed to del Ferro's son-in-law Hannibal Nave when del Ferro died in 1526.

On his deathbed, del Ferro revealed at least part of his secret, the solution to the "cube plus cosa equals number" problem, to his student, Fior.

1.4. Niccolo Tartaglia. Niccolo Fontana, known as Tartaglia, was born in Brescia in 1499 or 1500. His father was murdered when he was six, and plunged the family into total poverty.

Niccolo was nearly killed as a teenager when, in 1512, the French captured his home town and put it to the sword. The twelve year old Niccolo was dealt horrific facial sabre wounds by a French soldier that cut his jaw and palate. He was left for dead and even when his mother discovered that he was still alive she could not afford to pay for any medical help. However, his mother's tender care ensured that the youngster did survive, but in later life Niccolo always wore a beard to camouflage his disfiguring scars and he could only speak with difficulty, hence his nickname Tartaglia, or stammerer.

He moved to Venice in 1534. As a lowly mathematics teacher in Venice, Tartaglia gradually acquired a reputation as a promising mathematician by participating successfully in a large number of debates.

Fior began to boast that he was able to solve cubics and a challenge between him and Tartaglia was arranged in 1535. In fact Tartaglia had previously discovered how to solve one type of cubic equation, the "cube + squares equals number" type. For the contest between Tartaglia and Fior, each man was to submit thirty questions for the other to solve. Fior was supremely confident that his ability to solve cubics would be enough to defeat Tartaglia but Tartaglia submitted a variety of different questions, exposing Fior as an, at best, mediocre mathematician. Fior, on the other hand, offered Tartaglia thirty opportunities to solve the "cube plus cosa" problem, since he believed that he would be unable to solve this type, as in fact had been the case when the contest was set up. However, in the early hours of February 13, 1535, inspiration came to Tartaglia and he discovered the method to solve 'cube equal to numbers'. Tartaglia was then able to solve all thirty of Fior's problems in less than two hours. As Fior had made little headway with Tartaglia's questions, it was obvious to all who was the winner. Tartaglia didn't take his prize for winning from Fior, however, the honor of winning was enough.

1.5. Girolamo Cardano. Girolamo or Hieronimo Cardano's name was Hieronymus Cardanus in Latin and he is sometimes known by the English version of his name Jerome Cardan. He was the illegitimate child of a lawyer/mathematician, Fazio Cardano. He was a brilliant physician and mathematician who loved to gamble, and generally had a fascinating, though often tragic, life. Among other travails, he was kept out of the College of Physicians

because of his illegitimate birth; his wife died young; his favorite son was executed for the murder of his wife; his other son stole large sums of money from him; and he was jailed by the Inquisition.

In 1539,Cardan was a public lecturer of mathematics at the Piatti Foundation in Milan, and was aware of the problem of solving cubic equations; he had taken Pacioli at his word and assumed that, as Pacioli stated in the Suma published in 1494, solutions were impossible.

Cardan was greatly intrigued when he learned of the contest between Fior and Tartaglia, and he immediately set to work trying to discover Tartaglia's method for himself, but was unsuccessful. A few years later, in 1539, he contacted Tartaglia, through an intermediary, requesting that the method could be included in a book he was publishing that year. Tartaglia declined this opportunity, stating his intention to publish his formula in a book of his own that he was going to write at a later date. Cardan, accepting this, then asked to be shown the method, promising to keep it secret. Tartaglia, however, refused.

An incensed Cardan now wrote to Tartaglia directly, expressing his bitterness, challenging him to a debate but, at the same time, hinting that he had been discussing Tartaglia's brilliance with the governor of Milan, Alfonso d'Avalos, the Marchese del Vasto, who was one of Cardan's powerful patrons. On receipt of this letter, Tartaglia radically revised his attitude, realizing that acquaintance with the influential Milanese governor could be very rewarding and could provide a way out of the modest teacher's job he then held, and into a lucrative job at the Milanese court. He wrote back to Cardan in friendly terms, angling for an introduction to the Signor Marchese. Cardan was delighted at Tartaglia's new approach, and, inviting him to his house, assured Tartaglia that he would arrange a meeting with d'Avalos.

So, in March 1539, Tartaglia left Venice and travelled to Milan. To Tartaglia's dismay, the governor was temporarily absent from Milan but Cardan attended to his guest's every need and soon the conversation turned to the problem of cubic equations. Tartaglia, after much persuasion, agreed to tell Cardan his method, if Cardan would swear never to reveal it and furthermore, to only ever write it down in code so that on his death, nobody would discover the secret from his papers. The oath which Cardano swore is reportedly:

I swear to you, by God's holy Gospels, and as a true man of honor, not only never to publish your discoveries, if you teach me them, but I also promise you, and I pledge my faith as a true Christian, to note them down in code, so that after my death no one will be able to understand them.

Tartaglia divulged his formula in the form of a poem, to help protect the secret, should the paper fall into the wrong hands.

By the time he had reached Venice, Tartaglia was sure he had made a mistake in trusting Cardan and began to feel very angry that he had been induced to reveal his secret formula. When Cardan wrote to him in a friendly manner Tartaglia rebuffed his offer of continued friendship and mercilessly ridiculed his books on the merest trivialities.

1.6. Lodovico Ferrari. Lodovico Ferrari was sent, as a teenager, to be the servant of Cardano. However, when Cardano discovered that the boy could read and write, he made him his assistant, and quickly learned that Ferrari was quite talented. Ferrari became Cardano's mathematical apprentice.

Based on Tartaglia's formula, Cardan and Ferrari made remarkable progress finding proofs of all cases of the cubic and, even more impressively, solving the quartic equation. Tartaglia made no move to publish his formula despite the fact that, by now, it had become well known that such a method existed.

One of the first problems that Cardan hit was that the formula sometimes involved square roots of negative numbers even though the answer was a 'proper' number. In August 1539 Cardan wrote to Tartaglia:

I have sent to enquire after the solution to various problems for which you have given me no answer, one of which concerns the cube equal to an unknown plus a number. I have certainly grasped this rule, but when the cube of one-third of the coefficient of the unknown is greater in value than the square of one-half of the number, then, it appears, I cannot make it fit into the equation.

Indeed Cardan gives precisely the conditions here for the formula to involve square roots of negative numbers. Tartaglia by this time greatly regretted telling Cardan the method and tried to confuse him with his reply (although in fact Tartaglia, like Cardan, would not have understood the complex numbers now entering into mathematics):

... and thus I say in reply that you have not mastered the true way of solving problems of this kind, and indeed I would say that your methods are totally false.

Cardan and Ferrari travelled to Bologna in 1543 and learnt from Hannibal Nave that it had been del Ferro, not Tartaglia, who had been the first to solve the cubic equation. Cardan felt that although he had sworn not to reveal Tartaglia's method surely nothing prevented him from publishing del Ferro's formula. In 1545 Cardan published *Artis magnae sive de regulis algebraicis liber unus*, or *Ars magna*, as it is more commonly known, which contained solutions to both the cubic and quartic equations and all of the additional work he had completed on Tartaglia's formula. Del Ferro and Tartaglia are credited with their discoveries, as is Ferrari, and the story written down in the text.

It is to Cardan's credit that, although one could not expect him to understand complex numbers, he does present the first calculation with complex numbers in Ars Magna. Solving a particular cubic equation, he writes

Dismissing mental tortures, and multiplying $5 + \sqrt{-15}$ *by* $5 - \sqrt{-15}$*, we obtain* $25 - (-15)$*. Therefore the product is* 40 *... and thus far does arithmetical subtlety go, of which this, the extreme, is, as I have said, as subtle as it is useless.*

1.7. Rapheal Bombelli. In 1572, Rapheal Bombelli wrote his book *Algebra*, in which explicitly uses negative numbers and zero. Moreover, he shows how manipulating complex numbers can help arrive at real solutions to cubic equations, thus demonstrating that, although they may be subtle, they are far from useless.

2. Solution of Quadratic Equations

Some version of the quadratic formula has been available to most advanced cultures of the last three thousand years. Let us review its derivation.

A polynomial is *monic* if the leading coefficient is 1. Two equations are *equivalent* if they have the same solution set.

Let $ax^2 + bx + c = 0$ be a general quadratic equation; Our method of solution is known as completing the square. First we produce an equivalent monic equation, and then we introduce a new term to create the square of a linear polynomial:

$$\begin{aligned} ax^2 + bx + c = 0 &\Leftrightarrow x^2 + \frac{b}{a}x = -\frac{c}{a} \\ &\Leftrightarrow x^2 + \frac{b}{a}x + \left(\frac{b}{2a})^2 = (\frac{b}{2a}\right)^2 - \frac{c}{a} \\ &\Leftrightarrow \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2} \\ &\Leftrightarrow x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a} \\ &\Leftrightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \end{aligned}$$

Notice that the fourth equation can be rewritten as

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

Setting $y = x + \frac{b}{2a}$ and $n = \frac{4ac-b^2}{4a^2}$, rewrite this as

$$y^2 + n = 0.$$

This is *depressed quadratic equation*; the degree one term has be eliminated by the substitution $x \to y - \frac{b}{2a}$, which is known as a *linear change of variables*.

The *discriminant* of the quadratic equation is

$$\Delta = b^2 - 4ac;$$

this determines the number of real roots. There are three cases:

(a) if $b^2 - 4ac > 0$, there are two real roots;
(b) if $b^2 - 4ac = 0$, there is one real root;
(c) if $b^2 - 4ac < 0$, there are no real roots.

We point out here that negative numbers and there square roots were not accepted as actual solutions in antiquity. To some extent, this is justified, since the search for real solutions to such an equation was not significantly compromised by excluding these numbers. This situation changes upon consideration of cubic equations.

3. Depressing Cubic Equations

Let $f(x) = ax^3 + bx^2 + cx + d$ be a general cubic polynomial; we wish to solve $f(x) = 0$. If we can find one zero r of the polynomial $f(x)$, we can divide $(x - r)$ into $f(x)$ to obtain a quadratic polynomial, whose zeros can be found using the quadratic formula.

A *depressed cubic equation* of the type "cube plus cosa equals number" is an equation of the form

$$x^3 + mx = n.$$

We wish to take the general cubic equation $f(x) = 0$ and find a different depressed cubic equation whose solution will give us a solution to $f(x) = 0$.

Clearly we can divide by a to obtain a monic equation $x^3 + \frac{b}{a}x^2 + \frac{c}{a}x + \frac{d}{a} = 0$; in this way, we can assume that $a = 1$. We wish to find a linear change of variable that will produce an equation which lacks the quadratic term.

To discover how to do this, substitute $y - h$ for x and multiply:

$$\begin{aligned} ax^3 + bx^2 + cx + d &= a(y-h)^3 + b(y-h)^2 + c(y-h) + d \\ &= a(y^3 - hy^2 + h^2y - h^3) + b(y^2 - 2hy + h^2) + c(y-h) + d \\ &= ay^3 - 3ahy^2 + 3ah^2y - h^3 + by^2 - 2hy + h^2 + cy - ch + d \\ &= ay^3 + (b - 3ah)y^2 + (3ah^2 - 2h)y + (d - ch). \end{aligned}$$

Now we see that if we set $h = \frac{b}{3a}$, we obtain the desired result of eliminating the quadratic term.

Thus, to solve cubic equations, we may assume that the equation is given in the form $x^3 + mx = n$.

4. Solving the Depressed Cubic

Consider the depressed cubic equation

$$x^3 + mx = n.$$

The key idea of the solution method is to write x as a difference of two quantities, $x = t - u$. We wish to find appropriate quantities $t, u \in \mathbb{R}$ such that $r = t - u$, where r is a solution to this equation.

By the binomial theorem, we have

$$(t-u)^3 = t^3 - 3t^2u + 3tu^2 - u^3.$$

Thus

$$\begin{aligned} t^3 - u^3 &= (t-u)^3 + 3t^2u - 3tu^2 \\ &= (t-u)^3 + 3tu(t-u). \end{aligned}$$

Note that if $x = t - u$, $m = 3tu$, and $n = t^3 - u^3$, this equation becomes $x^3 + mx = n$. This reduces the problem to solving the system of equations

$$m = 3tu; \tag{1}$$

$$n = t^3 - u^3. \tag{2}$$

Equation (1) gives

$$u = \frac{m}{3t}.$$

Substitute this into equation (2) to obtain

$$t^3 - \frac{m^3}{27t^3} = n.$$

Subtract n from both sides and multiply through by t^3 to get

$$t^6 - nt^3 - \frac{m^3}{27} = 0.$$

Now the quadratic formula gives

$$\begin{aligned} t^3 &= \frac{n \pm \sqrt{n^2 + \frac{4m^3}{27}}}{2} \\ &= \frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{m^3}{27}}. \end{aligned}$$

Therefore,

$$t = \sqrt[3]{\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{m^3}{27}}}.$$

Now $u^3 = t^3 - n$, so

$$u = \sqrt[3]{-\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{m^3}{27}}}.$$

Finally, $x = t - u$, so

$$\boxed{x = \sqrt[3]{\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{m^3}{27}}} - \sqrt[3]{-\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{m^3}{27}}}}.$$

Upon examining the above equations, we see two $\pm$ signs in the expression for x, which may lead one to believe we have found four solutions to a cubic polynomial. However, to of them are the same.

Consider the solution to be in the form $x = \sqrt[3]{a \pm b} - \sqrt[3]{-a \pm b}$. The solution with two minus signs equals the solution with two negative signs, because factor out the negative sign gives

$$\sqrt[3]{a-b} - \sqrt[3]{-a-b} = (-1)\sqrt[3]{-a+b} - (-1)\sqrt[3]{a+b} = \sqrt[3]{a+b} - \sqrt[3]{-a+b}.$$

Memorizing this formula is unnecessary if we remember the technique. We let $x = t - u$, so that $x^3 + mx = n$ becomes $(t-u)^3 + 3tu(t-u) = t^3 - u^3$.

(1) Set $3tu = m$ and $t^3 - u^3 = n$.
(2) Solve the first equation for u to get $u = \frac{m}{3t}$.
(3) Plug this into the second equation to get $t^3 = n + (\frac{m}{3t})^3$.
(4) Multiply by t^3 to get $t^6 - nt^3 - (\frac{m}{3})^3 =$.
(5) Complete the square to get $t^3 = \frac{n}{2} \pm \Delta$.
(6) Use $u^3 = t^3 - n$ to get $u^3 = -\frac{n}{2} \pm \Delta$;
(7) Take cube root and set $x = t - u$.

Example 1. (A Typical Example)
Solve $x^3 + 15x = 22$.

Solution. Let $3tu = 15$ and $t^3 - u^3 = 22$. Then $u = \frac{5}{t}$, so $t^3 - \frac{125}{t^3} = 22$. Therefore $t^6 - 22t^3 - 125 = 0$, so by the quadratic formula,

$$t^3 = \frac{22 \pm \sqrt{484 + 500}}{2} = 11 \pm \sqrt{246}.$$

Then $u^3 = -11 \pm \sqrt{246}$, so

$$x = \sqrt[3]{11 \pm \sqrt{246}} - \sqrt[3]{-11 \pm \sqrt{246}}.$$

□

Example 2. (Cardano's Example)
Solve $x^3 + 6x = 20$.

Solution. Let $3tu = 6$ and $t^3 - u^3 = 20$. Then $u = \frac{2}{t}$, so $t^3 - \frac{8}{t^3} = 20$, so $t^6 - 20t^3 - 8 = 0$. By the quadratic formula,

$$t^3 = \frac{20 \pm \sqrt{400 + 32}}{2} = 10 \pm \sqrt{108}.$$

Taking the positive value for t, applying $u^3 = t^3 - 20$, and taking the appropriate cube roots, we have

$$x = t - u = \sqrt[3]{10 + \sqrt{108}} - \sqrt[3]{-10 + \sqrt{108}}.$$

Note that this is a real number.

A more modern solution starts by setting $f(x) = x^3 + 6x - 20$ and seeking solutions to $f(x) = 0$. We note that $f(2) = 2^3 + 6(2) - 20 = 0$, so $x = 2$ is a solution. By the Factor Theorem, $(x-2)$ divides $f(x)$, and dividing we find that

$$f(x) = (x-2)(x^2 + 2x + 10) = (x-2)(x - (-1+3i))(x - (-1-3i)).$$

Note that the only real zero of f is 2; thus, we have no choice but to conclude that

$$\sqrt[3]{10+\sqrt{108}}-\sqrt[3]{-10+\sqrt{108}}=2.$$

□

Example 3. (Bombelli's Example)
Solve $x^3-15x=4$.

Solution. Let $3tu=-15$ and $t^3-u^3=4$. Thus $u=-\frac{5}{t}$, so $t^3+\frac{125}{t^3}=4$. From this, $t^6-4t^3+125=0$, so, taking the positive square root, we have

$$t^3=\frac{4+\sqrt{16-500}}{2}=2+\sqrt{4-125}=2+\sqrt{-121}.$$

At this point, instead of asserting the irrelevance of the problem, Bombelli continues as if $\sqrt{-1}$ is a perfectly acceptable quantity. He notes that

$$(2+\sqrt{-1})^3=8+12\sqrt{-1}-6-\sqrt{-1}=2+11\sqrt{-1}=2+\sqrt{-121}.$$

Thus if $t=2+\sqrt{-1}$, then $t^3=2+\sqrt{-121}$. Continuing, we have $u=-2+\sqrt{-11}$, so

$$x=t-u=(2+\sqrt{-121})-(-2+\sqrt{-121})=4.$$

One verifies that 4 is indeed a solution to the original cubic equation; thus a real solution is attained by traversing through the realm of complex numbers. □

5. Depressing a Quartic Equation

The general quartic equation can be depressed in the same manner as the cubic. Consider

$$ax^4+bx^3+cx^2+dx+e=0.$$

We again want a linear change of variables that will eliminate the cubic term. Here, the substitution $x\mapsto(x-\frac{b}{4a}$ works.

Problem 1. Let

$$a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0=0.$$

Show that the substitution $x\mapsto(x-\frac{a_{n-1}}{na_n})$ eliminates the term of degree $(n-1)$.

6. Solving the Depressed Quartic

Consider the polynomial equation

$$x^4+px^2+qx+r=0.$$

Complete the square to obtain

$$(x^2+p)^2=px^2-qx-r+p^2.$$

Let $y\in\mathbb{R}$ and add $2y(x^2+p)+y^2$ to both sides to get

$$(*)\qquad (x^2+p+y)^2=(p+2y)x^2+(p^2-r+2py+y^2).$$

The right hand side becomes a quadratic in x we can choose y so that it is a perfect square; this is done by making the discriminant equal to zero:

$$q^2-4(p-r+2py+y^2)=0.$$

Rewrite this as

$$(q^2-4p^3+4pr)+(8r-16p^2)y-20py^2-8y^3=0.$$

This is a cubic in y, and can be solved. With this value for y, take the square root of both sides of (*) to obtain a quadratic in x, which can also be solved.

7. Graphs of Cubics

7.1. Backdrop. Cardano had several cases for solving cubics, placing the monomials on the appropriate side of the equation to create positive coefficients. As it turns out, the sign of the constant term plays no role in the computation, so the main cases were:

(a) $x^3 = mx + n$ (cube equals cosa plus number)
(b) $x^3 + mx = n$ (cube plus cosa equals number)

Analytic geometry on in cartesian coordinates had not been invented at the time, so Cardano had no idea that these two cases correspond to distinctly different geometric interpretations for the graph of the cubic. We investigate this in modern notation using Calculus.

7.2. The Leading Coefficient. Consider the generic cubic polynomial

$$f(x) = ax^3 + bx^2 + cx + d.$$

If $a > 0$, then $\lim_{x\to-\infty} f(x) = -\infty$, and $\lim_{x\to\infty} f(x) = \infty$; if $a < 0$, then $\lim_{x\to-\infty} f(x) = \infty$, and $\lim_{x\to\infty} f(x) = -\infty$ Thus by the Intermediate Value Theorem, f has at least one real zero. By the Factor Theorem, if $f(r) = 0$, then $f(x) = (x - r)q(x)$, where q is a quadratic polynomial, which we can find by polynomial division, and thus find the other two zeros (which may be real or complex) using the quadratic formula.

We wish to solve $f(x) = 0$, and we realize that if we divide through by a, the zeros of the resulting polynomials are the same as the original. Thus, without loss of generality, we assume $a = 1$.

7.3. The Square Coefficient. Consider the polynomial

$$f(x) = x^3 + bx^2 + cx + d.$$

Note that $\lim_{x\to-\infty} f(x) = -\infty$, and $\lim_{x\to\infty} f(x) = \infty$. Next we wish to discover the role of b in the graph of f.

By differentiation,

$$f'(x) = 3x^2 + 2bx + cand \quad f''(x) = 6x + 2b.$$

If $f''(x) = 0$, then $6x + 2b = 0$, so $x = -\frac{b}{3}$. Thus f has an inflection point at $(-\frac{b}{3}, f(-\frac{b}{3}))$. Also, $b = 0$ if and only if the inflection point of f lies on the y-axis. So, to eliminate the inflection point of f, we shift the graph of f right by $\frac{b}{3}$. The graph of $f(x - \frac{b}{3})$ is the graph of f shifted so that the inflection point is on the y-axis. If we find the zeros of $f(x - \frac{b}{3})$, we obtain the zeros of f subtracting $\frac{b}{3}$.

It turns out that

$$f(x - \frac{b}{3}) = x^3 - \frac{1}{3}(b^2 - 3c)x - \frac{1}{27}(b^3 - 3b^2 + 9bc - 27d).$$

Without loss of generality, we now assume that $b = 0$.

7.4. The Cosa Coefficient. Consider the polynomial

$$f(x) = x^3 + cx + d.$$

The graph of f has an inflection point on the y-axis. We wish to identify the role that c plays in the graph of f.

By differentiation,

$$f'(x) = 3x^2 + c.$$

Thus $f'(0)$ is the slope of the line tangent to the graph of f at its inflection point.

If $c = 0$, then f has a horizontal tangent at its inflection point.

If $c \geq 0$, then f is increasing and has no local minimum or maximum; in this case, f has exactly one x-intercept, and so, exactly one real zero and two complex zeros. This is the "cube + cosa = number" case.

If $c < 0$, we set $f'(x) = 0$ and obtain $x = \pm\sqrt{-\frac{c}{3}}$; thus f has a local maximum at $-\sqrt{-\frac{c}{3}}$ and a local minimum at $\sqrt{-\frac{c}{3}}$. This is the "cube = cosa + number" case. In this case, we sometimes have three distinct real zeros, and sometimes do not. We wish to discover a condition determining the number of real zeros.

7.5. The Constant Coefficient. Consider the polynomial

$$g(x) = x^3 + cx.$$

We have an excellent idea of its graph. It is an odd function, and thus has symmetry about the origin, and it has an inflection point at the origin. If $c \geq 0$, it is increasing, and if $c < 0$, it has zeros at $x = \pm\sqrt{c}$ with local extrema at $x = \pm\sqrt{-\frac{c}{3}}$.

If we shift this graph up by d (down if $d < 0$), the corresponding function is $f(x) = g(x) + d$.

Assume that $c < 0$. Let

$$h = g\left(\sqrt{-\frac{c}{3}}\right) = \frac{2c}{3}\sqrt{-\frac{c}{3}}.$$

This is the height (depth) of the local maximum (minimum). If we vertically shift the graph of g by less than h, we have three zeros; if we shift by h, we have one single zero and a double zero; if we shift by $d > h$, we have a unique zero. That is:

- **(a)** $|d| < |h| \Rightarrow$ three distinct real zeros
- **(b)** $|d| = |h| \Rightarrow$ one single real zero and one double real zero
- **(c)** $|d| > |h| \Rightarrow$ a unique real zero and two complex zeros

Now

$$|d| < |h| \Leftrightarrow d^2 < h^2 - \frac{4c^2}{9}(-\frac{c}{3}) = -\frac{4c^3}{27}$$
$$\Leftrightarrow \frac{d^2}{4} + \frac{c^3}{27}.$$

The *discriminant* of f is

$$D = \frac{d^2}{4} + \frac{c^3}{27}.$$

Note that if $c \geq 0$, then $D \geq 0$. Thus, for any c, we have

(a) $D < 0 \Rightarrow$ three distinct real zeros
(b) $D = 0 \Rightarrow$ one single real zero and one double real zero, or a triple zero
(c) $D > 0 \Rightarrow$ a unique real zero and two complex zeros

CHAPTER X

Ellipses

Abstract. Kepler realized that assigning elliptical orbits to the planets greatly simplified the description of their motion. Here we list basic facts about ellipses in modern notation.

1. Ellipses

Definition 1. An *ellipse* is the set of points in a plane such that the sum of the distances from the point to two given points, called *foci*, is a constant, called the *common sum*.

The midpoint between the foci is called the *center*. The line through the foci is called the *major axis*. The line perpendicular to the major axis through the center is called the *minor axis*. The points of intersection of the major axis with the ellipse are called *vertices*. The points of intersection of the minor axis with the ellipse are called *covertices*.

We also call the distance between the vertices the major axis, and half of it is the semimajor axis. Thus the semimajor axis is the distance from the center to a vertex.

2. Kepler's Laws of Planetary Motion

Law 1: The planets move in elliptical orbits with the sun at one vertex.

Law 2: The planets sweep out equal areas in equal amounts of time.

Law 3: The squares of the periods of the planets are proportional to the cubes of their semimajor axes.

3. Equations

Proposition 2. *Consider the ellipse with foci $(\pm c, 0)$, where $c > 0$, and common sum s. Then the center is $(0, 0)$, the major axis is $y = 0$, the minor axis is $x = 0$, the vertices are $(\pm a, 0)$, the covertices are $(0, \pm b)$, and the equation of the ellipse is*

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,}$$

where

$$\boxed{2a = s} \quad \text{and} \quad \boxed{c^2 = a^2 - b^2}.$$

Proof. The midpoint between the foci is clearly $(0, 0)$, so this is the center. Moreover, the line through $(\pm c, 0)$ is the x-axis, so its equation is $y = 0$, and the perpendicular line through the origin is the y-axis, which is $x = 0$.

Suppose that the equation of the ellipse is as stated. If (x, y) is on the intersection of the locus of this equation with the line $y = 0$, then $\frac{x^2}{a^2} = 1$, so $x = \pm a$; thus the vertices are $(\pm a, 0)$. Similarly, the covertices are $(0, \pm b)$.

Now from the definition of an ellipse, the distance from $(a, 0)$ to $(c, 0)$ plus the distance from $(a, 0)$ to $(-c, 0)$ equals s, that is,

$$s = (a - c) + (a + c) = 2a.$$

Moreover, the distance from $(0, b)$ to $(c, 0)$ plus the distance from $(0, b)$ to $(-c, 0)$ equals s. Thus

$$s = \sqrt{(c - 0)^2 + (0 - b)^2} + \sqrt{(-c - 0)^2 + (0 - b)^2} = 2\sqrt{c^2 + b^2}.$$

Since $s = 2a$, this gives $a = \sqrt{c^2 + b^2}$, so $a^2 = c^2 + b^2$, which we rewrite as $c^2 = a^2 - b^2$. It remains to derive the equation of the ellipse from the definition.

Let (x, y) be an arbitrary point on the ellipse; from the definition, we have

$$\sqrt{(x - c)^2 + (y - 0)^2} + \sqrt{(x - (-c))^2 + (y - 0)^2} = s.$$

Subtracting $\sqrt{(x + c)^2 + y^2}$ from both sides and squaring gives

$$(x - c)^2 + y^2 = s^2 + (x + c)^2 + y^2 - 2s\sqrt{(x + c)^2 + y^2}.$$

Rearranging this gives

$$2s\sqrt{(x + c)^2 + y^2} = s^2 + (x + c)^2 - (x - c)^2 = s^2 + 4cx.$$

Dividing by $2s$ and squaring again produces

$$x^2 + 2cx + c^2 + y^2 = \frac{s^2}{4} + 2cx + \frac{4c^2x^2}{s^2}.$$

Cancelling $2cx$ and using that $s^2 = -2a^2$ and $c^2 = a^2 - b^2$ leads us to

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

Adding $\frac{b^2x^2}{a^2} - x^2 - a^2 + b^2$ to both sides gives

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

Finally, dividing by b^2 gives

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

□

4. Eccentricity

The *eccentricity* of an ellipse is

$$\boxed{e = \frac{c}{a} = \frac{\text{distance between foci}}{\text{distance between vertices}}.}$$

Thus $c = ae$.

For an ellipse with $a > b$, we can compute b^2 in terms of a and e as

$$c^2 = a^2 - b^2 \quad \Rightarrow \quad b^2 = a^2 - c^2 = a^2 - a^2e^2 = a^2(1 - e^2).$$

The equation of the ellipse centered at the origin with semimajor axis a and eccentricity e is

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1.}$$

5. Area

Let's use calculus to compute the area of an ellipse with equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Solving for y gives

$$y = b\sqrt{1 - \frac{x^2}{a^2}} = \frac{b}{a}\sqrt{a^2 - x^2}.$$

Integrating from $-a$ to a gives the area of the upper half of the ellipse:

$$\int_{-a}^{a} \frac{b}{a}\sqrt{a^2 - x^2}\, dx = \frac{b}{a}\int_{-a}^{a} \sqrt{a^2 - x^2}\, dx = \frac{b}{a}\left[\frac{1}{2}\pi a^2\right].$$

We recognize the latter integral as that of a a semicircle of radius a, giving the stated value. So the area of the ellipse is double this:

$$\boxed{A = \pi ab}.$$

6. Reflectivity

Proposition 3. *Consider an ellipse with foci F_1 and F_2. Let P be a point on the ellipse and let L_0 be the line through P tangent to the ellipse. Let L_1 be the line through F_1 and P and let L_2 be the line through F_2 and P. Then the angle between L_0 and L_1 equals the angle between L_0 and L_2.*

Remark 4. This says that a wave emitted from one focus bounces off the surface and is transmitted to the other focus.

CHAPTER XI

Analytic Geometry

- René Descartes (French 1596-1650)
- Pierre de Fermat (French 1601-1665)
- Blaise Pascal (French 1623-1662)

René Descartes (French 1596-1650) was the author of mathematical studies and philosophical contemplations. He additionally undertook the construction of optical equipment.

He worked on the book *Le Monde* for four years, but as it neared completion, he forswore publication after hearing of the Galileo's experience with the Inquisition.

He published *A Discourse on the Method of Rightly Conducting the Reason and Seeking Truth in the Sciences* in 1637. This work included three appendices:

(1) *La Dioptrique*: a work on optics
(2) *Les Météores*: a work on meteorology
(3) *La Gèométrie*: a work on mathematics using coordinates to combine algebra and geometry, i.e., analytic geometry

Descartes on tangents.

CHAPTER XII

Power Series

Historical Background

Bonaventura Cavalieri (Italian 1598-1647)

- Introduced logarithms into Italy
- Wrote books on mathematics, optics, and astronomy
- Wrote *Geometria indivisibilibus*, published 1646, devoted to the *method of indivisibles* (parallel line segments in a plane region, of parallel region constituting a volume) producing *Cavalieri's principles.*

Cavalieri's principles are:

(a) If two planar pices are included between a pair of parallel lines, and if the lengths of the two segments cut by them on any line parallel to the including lines are always in a given ratio, then the areas of the two planar pieces are also in this ratio.

(b) If two solids are included between a pair of parallel palnes, and if the areas of the two sections cut by them on any plane parallel to the including planes are always in a given ratio, then the volumes of the two solids are also in this ratio.

John Wallis (English 1619-1703)

- Conics as degree two equations
- Extended methods of Descartes and Cavalieri
- Introduced symbol ∞
- Computed $\int_0^1 \sqrt{1-x^2}\,dx$ as an infinite product
- Computed arc lengths

Isaac Barrow (English 1630 - 1677)

- Proposed the "differential triangle", which is similar to Fermat's method but emphasizing slope.
- States and proved a version of the Fundamental Theorem of Calculus in *Lectiones optical et geometrical* (1670)

Isaac Newton (English 1642-1727)

- (1665) Early discoveries at Cambridge (1665-1666)
  - Generalized binomial theorem
  - Method of Fluxions (differential calculus)
  - Optics
  - Gravitation
- (1665) Compares decimal numbers to power series of variables with the analogy `series:algebra::decimal:arithmetic`
- (1668) Quadrature of the Hyperbola, published by Mercator in 1668
$$\int_0^x \frac{dt}{1+t} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} (= \log(1+t))$$
- (1669) Barrow resigned Lucasion chair, which was taken over by Newton. As such, Newton had to give lectures: "If he had an audience, they lasted 30 minutes; otherwise he spoke to the walls for 15 minutes and then left."
- (1671) Method of Fluxions written, but not published until 1736
- (1675) Corpuscular Theory of Light [Huygen's developed the wave theory]
- (1679) Verified gravitational equation $F \sim \frac{m_1 m_2}{r^2}$, and derived Kepler's laws of planetary motion from this.
- (1684) Halley convinced him to write ... wrote first book of *Philosohiae naturalis principia mathematica*
- (1687) Published at Halley's expense

Gottfried Leibniz (German 1646-1716)

- Diplomat, philosopher, lawyer
- Developed Calculus separately from Newton
- Notation: $\frac{dy}{dx}$, $\int y\,dx$ ($\int$ is an "S" from Latin word "summa", as a sum of Cavalieri's indivisibles)
- Introduced the word "function" and was the first to think in function terms.
- Distinction between algebraic and transcendental functions
- Preferred "closed form" to infinite series
- Also developed the theory of determinants

The views of Newton and Leibnitz differ on integration:

$$\int f(x)\,dx \quad \text{means} \quad \begin{cases} \text{Leibniz: find the closed form antiderivative} \\ \text{Newton: express } f(x) \text{ as a power series and lift each term} \end{cases}$$

1. Sequences

A *sequence* of real numbers is a function

$$a : \mathbb{N} \to \mathbb{R};$$

if $n \in \mathbb{N}$, we typically write a_n instead of $a(n)$. We denote the sequence $a : \mathbb{N} \to \mathbb{R}$ by (a_n).

Let (a_n) be a sequence and let $L \in \mathbb{R}$. We say that (a_n) *converges to* L if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$N < n \Rightarrow |a_n - L| < \epsilon.$$

If a sequence converges to a real number L, we say it is *convergent*, and we say that L is the *limit* of the sequence; we may write

$$L = \lim_{n \to \infty} a_n.$$

It is a fact that limits, when they exist, are unique.

If a sequence does not converge to a real number L, it is *divergent.*

One may form sums and products of sequences:

$$(a_n) + (b_n) = (a_n + b_n)$$
$$(a_n)(b_n) = (a_n b_n)$$

If (a_n) converges to L_1 and b_n converges to L_2, then $(a_n) + (b_n)$ converges to $L_1 + L_2$ and $(a_n)(b_n)$ converges to $L_1 L_2$.

If (a_n) is nonzero and converges to L, then

$$\lim_{n \to \infty} \frac{1}{a_n} = \frac{1}{L}.$$

Let (a_n) be a sequence. We say (a_n) is *increasing* if $a_m \leq a_n$ whenever $m \leq n$; we say that (a_n) is *decreasing* if $a_m \geq a_n$ whenever $m \leq n$; we say that (a_n) is *monotone* if it is either increasing or decreasing. We say that (a_n) is *bounded* if there exists a positive real number B such that $a_n \in [-B, B]$ for all $n \in \mathbb{N}$.

Fact 1. (Bounded Monotone Convergence Rule)
A bounded monotone sequence of real numbers converges.

Fact 2. (Squeeze Law)
If (a_n) and (b_n) both converge to L, and $a_n \leq c_n \leq b_n$ for all $n \in \mathbb{N}$, then (c_n) converges to L.

2. Series

Let (a_n) be a sequence. Then n^{th} *partial sum* of this series is

$$s_n = \sum_{i=0}^{n}.$$

A *series* is a sequence of the form (s_n), where s_n is the n^{th} partial sum of some sequence (a_n). Such a series may be denoted by $\sum a_n$.

A series $\sum a_n$ converges if the sequence of partial sums converges. In this case, we let $\sum_{n=0}^{\infty} a_n$ denote the limit of the series.

We say a series $\sum a_n$ *converges absolutely* if the associated series $\sum |a_n|$ converges. If a series converges absolutely, then it converges.

One may form sums and products of series:

$$\sum a_n + \sum b_n = \sum (a_n + b_n);$$
$$\sum a_n \sum b_n = \sum_{n=1}^{\infty} \left(\sum_{j=1}^{n} a_j b_{n-j} \right).$$

If $\sum a_n$ converges to S_1 and $\sum b_n$ converges to S_2, then $\sum a_n + \sum b_n$ converges to $S_1 + S_2$ and $\sum a_n \sum b_n$ converges to $S_1 S_2$.

Fact 3. (Limit Test)
If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

Reason. Set $s = \sum a_n$. Note that $a_n - s_n - s_{n-1}$, where $s_n = \sum_{i=1}^{n}$, so that $s = \lim s_n$. Now (s_{n-1}) is a sequence, whose limit is also clearly s. Thus

$$\lim a_n = \lim(s_n - s_{n-1}) = \lim s_n - \lim s_{n-1} = s - s = 0.$$

□

Fact 4. (Comparison Test)
Let $\sum c_n$ be a convergent series and let $\sum d_n$ be a divergent series.

(a) *If $0 \le a_n \le c_n$ for all $n \in \mathbb{N}$, then $\sum a_n$ converges.*
(b) *If $0 \le d_n \le b_n$ for all $n \in \mathbb{N}$, then $\sum b_n$ diverges.*

Fact 5. (Geometric Series Test)
Let $r \ge 0$.

(a) *If $r < 1$, then $\sum r^n$ converges to $\frac{1}{1-r}$.*
(b) *If $r \ge 1$, then $\sum r^n$ diverges.*

Reason. Note that $1-x^n = (1-x)(1+x+\cdots+x^{n-1})$; therefore $\frac{1-x^n}{1-x} = \sum_{i=0}^{n-1} x^i$. If $|x| < 1$, then $x^n \to 0$ as $n \to \infty$; thus

$$\begin{aligned}\sum_{i=0}^{\infty} x^n &= \lim_{n\to\infty} \sum_{i=0}^{n-1} x^i \\ &= \lim_{n\to\infty} \frac{1-x^n}{1-x} \\ &= \frac{1}{1-x}.\end{aligned}$$

□

Fact 6. (Alternating Series Test)
Let (a_n) be a decreasing sequence of nonnegative real numbers which converges to zero. Then $\sum(-1)^n a_n$ converges.

Reason. Note that $0 \leq s_2 \leq s_4 \leq s_6 \leq \cdots \leq a_1$. Thus (s_{2n}) is a bounded monotone sequence, and so it converges, say to s. Then $\lim s_{2n+1} = \lim s_{2n} + \lim a_{2n+1} = s + 0 = s$. □

Fact 7. (Ratio Test)
Let (a_n) be a sequence of positive real numbers such that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = L.$$

Then $\sum a_n$ converges if $L < 1$ and $\sum a_n$ diverges if $L > 1$.

Reason. Suppose $0 < L < 1$. Select r such that $0 < L < r < 1$. Let N be so large that

$$\left|\frac{a_{n+1}}{a_n}\right| < r \quad \text{for} \quad n \geq N.$$

Then $|a_{n+1}| < r|a_n|$, for $n \geq N$.

In particular, $|a_{N+1}| < r|a_N|$, $|a_{N+2}| < r|a_{N+1}| < r^2|a_N|$, and in general, $|a_{N+k}| < r^k|a_N|$. Now

$$\sum_{k=1}^{\infty} |a_n| < \sum_{k=1}^{\infty} |a_N| r^k,$$

which converges. □

Fact 8. (Root Test)
Let (a_n) be a sequence of positive real numbers such that

$$\lim_{n\to\infty} \sqrt{n} a_n = L.$$

Then $\sum a_n$ converges if $L < 1$ and $\sum a_n$ diverges if $L > 1$.

3. Power Series

A *power series* centered at $x_0 \in A$, where $A \subset \mathbb{R}$, is a function

$$f : A \to \mathbb{R}$$

which can be expressed in the form

$$f(x) = \sum_{i=0}^{\infty} a_n (x - x_0)^n.$$

Here, A is the set of points $x \in A$ where $f(x)$ converges. First we want to understand the set A. If we say $R \in [0, \infty]$, we mean that R is either a nonnegative real number or $R = \infty$.

Fact 9. *Let $f(x) = \sum a_n (x - x_0)^n$ be a power series. Then there exists a number $R \in [0, \infty]$ such that*

(a) *$f(x)$ converges absolutely if $|x - x_0| < R$;*
(b) *$f(x)$ diverges if $|x - x_0| > R$.*

This number R is called the radius of convergence *of f.*

We may compute the radius of convergence using our knowledge of series; in particular, the ratio test is useful. Suppose that

$$\lim_{n \to \infty} \frac{|a_{n+1}|}{|a_n|} = L.$$

Let $r = x - x_0$. Then

$$\lim_{n \to \infty} \frac{|a_{n+1}(x - x_0)^{n+1}|}{|a_n (x - x_0)^n|} = \lim_{n \to \infty} |\frac{a_{n+1} r}{a_n}| = rL.$$

Now $f(x) = \sum a_n (x - x_0)^n$ converges at the point x if $rL < 1$, which happens if $r < \frac{1}{L}$. On the other hand, if $r > \frac{1}{L}$, then $f(x)$ diverges. Thus the radius of convergence is $R = \frac{1}{L}$, i.e.,

$$R = \lim_{n \to \infty} \frac{|a_n|}{|a_{n+1}|}.$$

Similarly, we can use the root test to derive the formula

$$R = \lim_{n \to \infty} \frac{1}{\sqrt[n]{a_n}}.$$

Let $f(x) = \sum a_n (x - x_0)^n$ be a power series and let R be its radius of convergence. The *interval of convergence* of $f(x)$ the open interval $I = (x_0 - R, x_0 + R)$; if $R = \infty$, we take this to mean $I = \mathbb{R}$.

4. Power Series Algebra

We have defined power series as functions, and they behave very much like polynomial functions in a couple of ways.

Two functions are equal if and only if they have the same domain and range and take on the same value at every point in the domain. The following gives a useful condition for two power series to be equal; this condition is directly analogous to the condition for polynomial functions.

Fact 10. *Let $f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n$ and $g(x) = \sum_{n=0}^{\infty} b_n(x - x_0)^n$ be two power series centered at x_0. Then $f = g$ as functions if and only if $a_n = b_n$ for every $n \in \mathbb{N}$.*

The sum and product of functions is defined pointwise: $(f + g)(x) = f(x) + g(x)$, and $(fg)(x) = f(x)g(x)$. In the polynomial case, these can be obtained by distribution and reassociation. This remains true for power series.

Fact 11. *Let $f(x) = \sum_{n=0}^{\infty} a_n(x - x_0)^n$ and $g(x) = \sum_{n=0}^{\infty} b_n(x - x_0)^n$ be two power series centered at x_0. Then $f + g$ and fg are power series given by*

$$(f + g)(x) = \sum_{n=0}^{\infty} (a_n + b_n)(x - x_0)^n$$

and

$$(fg)(x) = \sum_{n=0}^{\infty} \left[\sum_{i=0}^{n} a_i b_{n-i} \right] (x - x_0)^n.$$

5. Shifting the Index of a Power Series

Let $k \in \mathbb{Z}$ and consider the infinite sum

$$\sum_{n=k}^{\infty} a_n(x - x_0)^n.$$

If $k < 0$, then this is not a power series. However, if $k > 0$, we consider this to be the power series by understanding that $a_i = 0$ for $i = 0, \ldots, k - 1$.

It is sometimes convenient to shift the index of a power series. The following is a formula for doing so:

$$\sum_{n=k}^{\infty} a_n(x - x_0)^n = \sum_{n=0}^{\infty} a_{n-k}(x - x_0)^{n-k}.$$

6. Differentiation of Power Series

It seems reasonable one may pass the differentiation operator inside the infinite sum:

$$\begin{aligned}\frac{d}{dx}\sum_{n=0}^{\infty} a_n(x-x_0)^n &= \sum_{n=0}^{\infty}\frac{d}{dx}(a_n(x-x_0)^n)\\ &= \sum_{n=0}^{\infty} n a_n(x-x_0)^{n-1}.\end{aligned}$$

This is indeed the case.

Fact 12. *Let $f(x)=\sum a_n(x-x_0)^n$ be a power series. Then f is differentiable in its radius of convergence, and*

$$f'(x)=\sum_{n=1}^{\infty} n a_n(x-x_0)^{n-1}.$$

Let $f(x)=\sum a_n(x-x_0)^n$ be a power series. We know attempt to find a formula which relates the derivatives of f to the coefficients a_n.

Note that for any power series, if we evaluate it at its center, we pick out the first coefficient because all of the other terms vanish at the center. By successively differentiating the power series, we shift the coefficients to the left. At each stage we write the first few terms to see how this goes.

Start with

$$f(x)=a_0+a_1(x-x_0)+a_2(x-x_0)^2+a_3(x-x_0)^3+\ldots;$$

thus $f(x_0)=a_0$, since all of the other terms in the series are of the form $a_n(x-x_0)^n$ and so they vanish at x_0.

Now $f'(x)$ is the power series

$$f'(x)=a_1+\frac{a_2}{2}(x-x_0)+\frac{a_3}{3}(x-x_0)^2+\frac{a_4}{4}(x-x_0)^3+\ldots;$$

by plugging in x_0, we pick off the constant coefficient; this time, we get $f'(x_0)=a_1$.

Differentiating again shifts the coefficients to the left to get

$$f''(x)=\frac{a_2}{2}+\frac{a_3}{2\cdot 3}(x-x_0)+\frac{a_4}{3\cdot 4}(x-x_0)^2+\frac{a_5}{4\cdot 5}(x-x_0)^3+\ldots;$$

thus $f''(x_0)=\frac{a_2}{2}$.

One more time gives

$$f'''(x)=\frac{a_3}{2\cdot 3}+\frac{a_4}{2\cdot 3\cdot 4}(x-x_0)+\frac{a_5}{3\cdot 4\cdot 5}(x-x_0)^2+\frac{a_6}{4\cdot 5\cdot 6}(x-x_0)^3+\ldots;$$

thus $f'''(x)=\frac{a_3}{6}$.

We now see the pattern; by the n^{th} differentiation, the n^{th} coefficient has moved into the constant coefficient position, but has been divided by $n!$ along the way. This gives us our main formula regarding power series.

Fact 13. *Let $f(x)=\sum a_n(x-x_0)^n$ be a power series with positive radius of convergence. Then*

$$a_n=\frac{f^{(n)}(x_0)}{n!}.$$

7. Taylor Series and Analytic Functions

Let $I \subset \mathbb{R}$ be an open interval and let $g : I \to \mathbb{R}$ be a smooth function on I. Let $x_0 \in I$.

The *Taylor series* of f expanded around x_0 There is a natural power series associated to the function g and the point x_0, called the *Taylor series* of f expanded around x_0, and defined by

$$f(x) = \sum a_n (x - x_0)^n,$$

where

$$a_n = \frac{f^{(n)}(x_0)}{n!}.$$

Note that if g is already a power series, it is equal to the associated power series around any point $x_0 \in I$.

We say that g is *analytic at* x_0 if there exists a sequence (a_n) of real numbers and a real number $R > 0$ such that for every $x \in I \cap (x_0 - R, x_0 + R)$, the power series

$$f(x) = \sum a_n (x - x_0)^n$$

converges, and $f(x) = g(x)$. We say that g is *analytic* if it is analytic at every point in I.

We see that g is analytic when it is equal to its Taylor series expansion around any point, and that the constant R above can be taken to be the radius of convergence of the Taylor series.

Fact 14. *Let $f : I \to \mathbb{R}$ be analytic at $x_0 \in I$ with radius of convergence R. Let $x_1 \in I \cap (x_0 - R, x_0 + R)$. Then f is analytic at x_1, with radius of convergence greater than or equal to* $\min\{x_1 - x_0 + R, x_0 + R - x_1\}$.

Let $f : I \to \mathbb{R}$ and $g : I \to R$ be analytic, and let $c \in \mathbb{R}$ be a constant. Then $f + g : I \to \mathbb{R}$, $cf : I \to \mathbb{R}$, and $fg : I \to \mathbb{R}$ are also analytic. Quotients of analytic functions are analytic in their domain of definition (with one caveat we will see later). If f and g are expanded around the same point $x_0 \in I$, the radius of convergence of these derived functions is at least as large as the minimum radius of convergence between f and g.

Let $\mathcal{A}(I) = \{f : I \to \mathbb{R} \mid f \text{ is analytic } \}$. Then $\mathcal{A}(I)$ is a vector space over $\mathbb{R}$.

Let $f(x) = \sum a_n (x - x_0)^n$ be analytic. We say that $f(x)$ is *entire* if the radius of convergence of f around x_0 is infinite. When this is the case, the radius of convergence of f expanded around any real number is still infinite.

The following functions are entire: constants, polynomials, exp, sin, and cos.

8. Standard Examples

Example 15. Find the Taylor expansion for $f(x) = \exp(x)$ around 0 and its radius of convergence.

Solution. All derivatives of f are the same. Thus the coefficients are simply

$$a_n = \frac{f^n(0)}{n!} = \frac{\exp(0)}{n!} = \frac{1}{n!}.$$

Thus

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The radius of convergence is

$$R = \lim_{n\to\infty} \frac{1/n!}{1/(n+1)!} = \lim_{n\to\infty} n+1 = \infty.$$

Thus exp is entire. □

Example 16. The Taylor expansion of $\sin(x)$ around 0 is given by

$$\sin(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots.$$

Thus sin is entire by the alternating series test.

Example 17. The Taylor expansion of $\cos(x)$ around 0 is given by

$$\cos(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots.$$

Thus cos is entire by the alternating series test.

Example 18. Find the Taylor expansion for $f(x) = \tan x$ around 0 and its radius of convergence.

Solution. First we take derivatives:

$$f'(x) = \sec^2 x; \quad f''(x) = 2\sec^2 x \tan x; \quad f'''(x) = 4\sec^2 x \tan^2 x + \sec^4 x.$$

Now we evaluate at 0:

$$f(0) = 0; \quad f'(0) = 1; \quad f''(0) = 0; \quad f'''(0) = 1.$$

□

Example 19. The Taylor expansion of $\log(1+x)$ around 0 is given by

$$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}.$$

Its radius of convergence is 1.

Example 20. Compute the Taylor expansion of $f(x) = \frac{1}{1+x}$ around $x_0 = 0$ and find its radius of convergence.

Solution. First, we differentiate until we begin to see a pattern. Then we plug in 0.

$$\begin{aligned} f(x) &= \frac{1}{1+x} & f(0) &= 1 = 0! \\ f'(x) &= \frac{-1}{(1+x)^2} & f'(0) &= -1 = -1! \\ f''(x) &= \frac{2}{(1+x)^3} & f''(0) &= 2 = 2! \\ f'''(x) &= \frac{-6}{(1+x)^4} & f'''(0) &= -6 = -3! \\ f^{iv}(x) &= \frac{24}{(1+x)^5} & f^{iv}(0) &= 24 = 4! \end{aligned}$$

We see that $f^{(n)}(0) = (-1)^n n!$. Then $a_n = (-1)^n$, and

$$f(x) = \sum_{n=0}^{\infty} (-1)^n x^n.$$

There is an easier way to do this by using the geometric series. Let $r = -x$; then

$$f(x) = \frac{1}{1-r} = \sum_{n=0}^{\infty} r^n = \sum_{n=0}^{\infty} (-1)^n x^n.$$

The radius of convergence is 1. □

We see that, in this example, the radius of convergence centered at x_0 is the distance from x_0 to the nearest point of discontinuity.

Example 21. Compute the Taylor expansion of $g(x) = \frac{1}{1+x^2}$ around $x_0 = 0$ and find its radius of convergence.

Solution. Note that $g(x) = f(x^2)$, where $f(x) = \frac{1}{1+x}$. Then

$$g(x) = \sum_{n=0}^{\infty} (-1)^n x^{2n}.$$

This is a power series with the coefficients of the odd terms all equal to zero. Its radius of convergence is still equal to 1. □

In this example, the function $g(x)$ is continuous and analytic at every point $x \in \mathbb{R}$. Then why does it have a finite radius of convergence? We answer this after one more example.

Example 22. Compute the Taylor expansion of $f(x) = \arctan(x)$ around $x_0 = 0$ and find its radius of convergence.

Solution. Let $f(x) = \arctan(x)$. Then $f'(x) = \frac{1}{1+x^2}$; view this as a geometric series. This produces

$$\begin{aligned} f'(x) &= \frac{1}{1+x^2} \\ &= \frac{1}{1-(-x^2)} \\ &= \sum_{n=0}^{\infty} (-x^2)^n \\ &= \sum_{n=0}^{\infty} (-1)^n x^{2n} \\ &= 1 - x^2 + x^4 - x^6 + x^8 + \cdots . \end{aligned}$$

Now

$$\begin{aligned} f(x) &= \int \frac{1}{1+x^2}\, dx \\ &= \int \left(\sum_{n=0}^{\infty} (-1)^n x^{2n} \right) dx \\ &= \sum_{n=0}^{\infty} (-1)^n \int x^{2n}\, dx \\ &= \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \\ &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots . \end{aligned}$$

□

9. Binomial Theorem

Power series are a generalization of polynomials. The extent to which this is true is illuminated by the generalized binomial theorem, discovered by Isaac Newton in the seventeenth century.

Let $n \in \mathbb{N}$, and define $\binom{n}{i}$ to be the number of possible ways to choose a set of i things from a set of n things. We see that there are n choices for the first thing, $n-1$ choices for the second, and so forth, until finally there are $n-i+1$ choices for the i^{th} thing. Thus there are $n(n-1)\cdots(n-i) = \frac{n!}{(n-i)!}$ possibilities for choosing i things, *in a certain order*. There are $i!$ possible different orders for the same set of i things, so altogether we have

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

These numbers are exactly those which are produces via Pascal's Triangle, and are called the *binomial coefficients*. This name comes from the binomial theorem for positive integers, which we state as

$$(x+1)^n = \sum_{i=0}^{n} \binom{n}{i} x^i.$$

This may be thought of as follows: multiplying $x+1$ by itself n times using distribution involves 2^n multiplications of n things, either x or 1 from each of the n copies of the $(x+1)$'s that are being multiplied. Each such multiplication involves choosing either x or 1 from each binomial $(x+1)$. There are $\binom{n}{i}$ different ways of selecting i x's and $(n-i)$ 1's. When we collect like terms, the coefficient of x^i is the number of x^i's occurring in the sum; this is $\binom{n}{i}$.

Suppose that $i > n$; there are zero ways of choosing a set of i items from of a set of n items, so the natural definition in this case is $\binom{n}{i} = 0$. In this case, we may write $(x+1)^n$ as a power series:

$$(x+1)^n = \sum_{i=0}^{\infty} \binom{n}{i} x^i,$$

because $\binom{n}{i} = 0$ for $i > n$.

Newton saw this, and generalized it in the following fashion, which we explain in modern notation.

Let $\alpha \in \mathbb{R}$, and consider the function $f(x) = (x+1)^\alpha$, so that

$$\begin{aligned} f(x) &= (x+1)^\alpha, \\ f'(x) &= \alpha(x+1)^{\alpha-1}, \\ f''(x) &= \alpha(\alpha-1)(x+1)^{\alpha-2}, \\ f'''(x) &= \alpha(\alpha-1)(\alpha-2)(x+1)^{\alpha-3}, \end{aligned}$$

and so forth. Generalizing the binomial coefficients, for $i \in \mathbb{N}$ define

$$\binom{\alpha}{i} = \frac{\prod_{j=0}^{i-1}(\alpha-j)}{i!}.$$

Here we use the convention that the empty product is 1, so $\binom{\alpha}{0} = 1$. Then we see that

$$f^{(i)}(x) = i!\binom{\alpha}{i-1}(x+1)^{\alpha-i}.$$

Evaluating this at $x_0 = 0$, we have

$$f^{(i)}(0) = i!\binom{\alpha}{i-1}.$$

Thus the i^{th} coefficient of the Taylor expansion of f around $x_0 = 0$ is

$$a_i = \frac{f^{(i)}(0)}{i!} = \binom{\alpha}{i},$$

so the Taylor series of $f(x)$ is

$$f(x) = \sum_{i=0}^{\infty} \binom{\alpha}{i} x^i;$$

this is known as the *binomial series*. The radius of convergence of the binomial series is

$$R = \lim_{i\to\infty} \left| \frac{\alpha(\alpha-1)\cdots(\alpha-i+1)/i!}{\alpha(\alpha-1)\cdots(\alpha-i)/(i+1)!} \right| = \lim_{i\to\infty} \left| \frac{i+1}{\alpha-i} \right| = \lim_{i\to\infty} \left| \frac{1+\frac{1}{i}}{1-\frac{\alpha}{i}} \right| = 1.$$

For example, let $f(x) = \sqrt{1-x} = (1-x)^{\frac{1}{2}}$. Then

$$\begin{aligned}
\sqrt{1-x} &= \sum_{i=0}^{\infty} \binom{1/2}{i} (-x)^i \\
&= 1 + \frac{1}{2}(-x) + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\frac{(-x)^2}{2!} + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{(-x)^3}{3!} \\
&\quad + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\frac{(-x)^4}{4!} \\
&\quad + \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\left(-\frac{7}{2}\right)\frac{(-x)^5}{5!} + \cdots \\
&= 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3 - \frac{5}{128}x^4 - \frac{7}{256}x^5 + \cdots
\end{aligned}$$

Newton used this to estimate $\sqrt{3}$, using that

$$\sqrt{3} = \sqrt{4-1} = 2\sqrt{1-\frac{1}{4}}.$$

Letting $x = \frac{1}{4}$, we have

$$\begin{aligned}
\sqrt{3} &\approx 2\left(1 - \frac{1}{2}\cdot\frac{1}{4} - \frac{1}{8}\cdot\frac{1}{16} - \frac{1}{16}\cdot\frac{1}{64} - \frac{5}{128}\cdot\frac{1}{256} - \frac{7}{256}\cdot\frac{1}{1024}\right) \\
&= 2 - \frac{1}{4} - \frac{1}{64} - \frac{1}{512} - \frac{5}{16384} - \frac{7}{131072} \\
&\approx 1.732063293.
\end{aligned}$$

Let $s = 1.732063293$; to nine decimal places, the actual value is $\sqrt{3} \approx 1.732050808$. To get more accuracy, Newton could have just used a few more terms.

10. Newton's Approximation for π

Consider the function $f(x) = \sqrt{x - x^2}$. The graph of this function is the upper half of a circle of radius one half centered at the point $(\frac{1}{2}, 0)$. Compute the area under the curve between $x = 0$ and $x = \frac{1}{4}$ in two ways; using the method of "fluxions", and then using geometry.

10.1. Area by Fluxions. The method of fluxions, expressed in modern language, consists of expanding functions into their Taylor series, the differentiating or integrating term by term. The area of which speak is

$$\begin{aligned}
\int_0^{\frac{1}{4}} f(x)\,dx &= \int \sqrt{x}\sqrt{1-x}\,dx\Big|_{\frac{1}{4}} \\
&\approx \int \sqrt{x}\left(1 - \frac{1}{2}x - \frac{1}{8}x^2 - \frac{1}{16}x^3 - \frac{5}{128}x^4 - \frac{7}{256}x^5\right)dx\Big|_{\frac{1}{4}} \\
&= \int \left(x^{1/2} - \frac{1}{2}x^{3/2} - \frac{1}{8}x^{5/2} - \frac{1}{16}x^{7/2} - \frac{5}{128}x^{9/2} - \frac{7}{256}x^{11/2}\right)dx\Big|_{\frac{1}{4}} \\
&= \frac{2}{3}x^{3/2} - \frac{1}{5}x^{5/2} - \frac{1}{28}x^{7/2} - \frac{1}{72}x^{9/2} - \frac{5}{704}x^{11/2} - \frac{7}{1664}x^{13/2}\Big|_{\frac{1}{4}} \\
&= \frac{2}{3}\left(\frac{1}{2}\right)^3 - \frac{1}{5}\left(\frac{1}{2}\right)^5 - \frac{1}{28}\left(\frac{1}{2}\right)^7 - \frac{1}{72}\left(\frac{1}{2}\right)^9 - \frac{5}{704}\left(\frac{1}{2}\right)^{11} - \frac{7}{1664}\left(\frac{1}{2}\right)^{13} \\
&= \frac{1}{12} - \frac{1}{1670} - \frac{1}{3584} - \frac{1}{36864} - \frac{5}{1441792} - \frac{7}{13631488} \\
&\approx 0.076773207.
\end{aligned}$$

Let $a = 0.076773207$; this is our approximation for the area of the region being considered.

10.2. Area by Geometry. Let $O = (0, 0)$, $A = (\frac{1}{4}, 0)$, $B = (\frac{1}{2}, 0)$, and $C = (\frac{1}{4}, \frac{\sqrt{3}}{4})$. Then C is on the semicircle $y = \sqrt{x - x^2}$. The sector OBC is one sixth of this circle if radius $\frac{1}{2}$, so its area is $\frac{\pi}{24}$. The triangle ABC has area $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{\sqrt{3}}{4}$; thus the area of the sector is $\frac{\pi}{24} - \frac{\sqrt{3}}{32}$. This is approximately equal to a, so

$$\pi \approx 24(a + \frac{\sqrt{3}}{32}) \approx 24(a + \frac{s}{32}),$$

where $s = 1.732063293$ is the approximation for $\sqrt{3}$ we obtained in the last section. Thus

$$\pi \approx 3.141604438.$$

Newton actually carried this approximation out using nine terms of the binomial expansion, and obtained an estimate for π which was accurate to the seventh decimal place.

11. Analytic Functions and Complex Numbers

Why do some functions have a finite radius of convergence? For example, we know that $\tan x$ is not defined wherever $\cos x = 0$, for example at $x = \frac{pi}{2}$, so if we expand $\tan x$ around $x_0 = 0$, we are bound to see that the radius of convergence is no bigger than $\frac{\pi}{2}$; on the other hand, since $\sin x$ and $\cos x$ are entire and $\cos x$ is nonzero in $I = (-\frac{\pi}{2}, \frac{\pi}{2})$, we expect that $\tan x$ is analytic in I so the radius of convergence of the expansion around 0 should be exactly $\frac{\pi}{2}$, which turns out to be the case.

However, this doesn't explain the radius of convergence of the function $f(x) = \frac{x}{1+x^2}$, which is analytic in the interval $I = (-1, 1)$, but when expanded around zero has a radius of convergence of only 1. The numerator and denominator are analytic and the denominator is nonvanishing for all real numbers x; why isn't f analytic? To understand this, we must expand our vision to the complex plane.

Our entire theory of sequences, series, power series, and Taylor series generalizes to use of complex numbers. A complex power series has a *disk of convergence*; if

$$f(z) = \sum a_n (z - z_0)^n,$$

where $a_n, z_0 \in \mathbb{C}$, then f converges in a disk around z_0 of radius R, where R is the radius of convergence as computed above (the absolute value of a complex number is its modulus).

The answer to our question is: the radius of convergence is the distance to the nearest *nonremovable complex singularity*. Let us examine what this means.

12. Laurent Series

Let $I \subset \mathbb{R}$ be an open interval. Let $x_0 \in I$ and let $A = I \setminus \{x_0\}$.

A (inessential) *Laurent series* at x_0 is a function $g : A \to \mathbb{R}$ such that there exists an integer $k \in \mathbb{Z}$ and real number $a_k, a_{k+1}, \cdots \in \mathbb{R}$ such that

$$g(x) = \sum_{n=k}^{\infty} a_k (x - x_0)^k.$$

If $k \geq 0$, a Laurent series is a power series.

Let $f : I \to \mathbb{R}$ be analytic. We attempt to find a Laurent series for $\frac{1}{f}$ at x_0. In particular, we try to find the number of negatively indexed coefficients in the inverse of f.

If we let $a_n = \frac{f^{(n)}}{n!}$, then

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n.$$

We seek a function

$$g(x) = \sum_{n=k}^{\infty} b_n (x - x_0)^n,$$

where $k \leq 0$ and $b_k \neq 0$, such that $fg(x) = 1$ for every $x \in I$. Let c_n be the n^{th} term in the product; the lowest possible value for n is k. Then $c_n = \sum_{i-j=0} a_i b_j$;

when we multiply these series, we should get

$$\begin{aligned} c_k &= a_0 b_k \\ c_{k+1} &= a_0 b_{k+1} + a_1 b_k \\ c_{k+2} &= a_0 b_{k+2} + a_1 b_{k+1} + a_2 b_{k+3} \\ &\vdots \\ c_{-1} &= a_0 b_{-1} + \cdots + a_{k-1} b_k \\ c_0 &= a_0 b_0 + \cdots + a_k b_k \\ c_1 &= a_0 b_1 + \cdots + a_{k+1} b_k \end{aligned}$$

We want $c_0 = 1$ and all other $c_n = 0$. Then we better have $a_0 = 0$ (consider c_k), whence $a_1 = 0$ (considering c_{k+1}), and so forth up to a_{k-1}. The first a_n which is not equal to zero is at $n = k$.

13. Singularities

Let $I \subset \mathbb{R}$ be an open interval. Let $x_0 \in I$ and let $A = I \setminus \{x_0\}$. Let $g : A \to \mathbb{R}$ be analytic on A.

We say that g is *meromorphic* at x_0 if we may write g as an inessential Laurent series centered at x_0. We say that x_0 is a *singularity* of g.

Let $g : I \to \mathbb{R}$ be meromorphic at x_0. We say that the singularity at x_0 is *removable* if $\lim_{x \to x_0} g(x)$ exists; in this case, we may define

$$f(x) = \begin{cases} g(x) & \text{if } x \neq x_0; \\ \lim_{x \to x_0} f(x) & \text{if } x = x_0. \end{cases}$$

Then $f(x)$ is analytic at x_0; we think of f and g as interchangeable, and can write f as a power series around x_0.

We say that g has a *zero of order* n at x_0 if n is smallest integer such that the n^{th} coefficient of the Laurent expansion of f is nonzero. Equivalently, this is the maximum positive integer n such that $\frac{g(x)}{x^n}$ has a removable singularity at x_0.

We say that g has a *pole of order* n at x_0 if n is the minimum number of negatively indexed terms in the Laurent expansion of g. Equivalently, this is the maximum positive integer n such that $(x - x_0)^n g(x)$ has a removable singularity at x_0.

Note that g has a pole of order n at x_0 if and only if g has a zero of order $-n$ at x_0.

If f has a zero of order n at x_0 and g has a pole of order n at x_0, then fg has a removable singularity at x_0, and $fg(x_0) \neq 0$; equivalently, fg has a zero of order 0 at x_0.

CHAPTER XIII

Complex Numbers

Historical Background

Reference:
`http://math.fullerton.edu/mathews/n2003/ComplexNumberOrigin.html`.

Rafael Bombelli (Italian 1526 - 1572)
Recall that Cardano, in attempting to solve the cube equals cosa plus number case $x^3 = mx+n$, arrived at a negative sign under the radical. Tartaglia rebuked him, claiming that his methods were "totally false". Cardano, in attempting to go forward with this, eventually claimed that such considerations were "as subtle as they are useless".

However, in his 1572 treatise L'Algebra, Rafael Bombelli showed that roots of negative numbers have great utility indeed. Consider the depressed cubic $x^3 = 15x + 4$. Applying the method of Tartaglia and Cardano, we set $m = -15$ and $n = 4$. If $x = t - u$, then $3tu = m = -15$ and $t^3 - u^3 = n = 4$, so that $u^3 = -\frac{125}{t^3}$, and $t^3 + \frac{125}{t^3} = 4$. Then $t^6 - 4t^3 - 125 = 0$, and by the quadratic formula, $t^3 = 2 + \sqrt{-121} = 2 + 11\sqrt{-1}$, whence $u^3 = -2 + 11\sqrt{-1}$, and $x = \sqrt[3]{2 + 11\sqrt{-1}} - \sqrt[3]{-2 + 11\sqrt{-1}}$.

Now Bombelli, undeterred by the negative sign under the radical, wished to find a number whose cube was $2+11\sqrt{-1}$. Having a "wild thought", he assumed that such a number would be of the form $a + b\sqrt{-1}$. This produces

$$(a + b\sqrt{-1})^3 = (a^3 - 3ab^2) + (3a^2b - b^3)\sqrt{-1} = 2 + 11\sqrt{-1},$$

from which we conclude that $a^3 - 3ab^2 = 2$ and $3a^2b - b^3 = 11$. The first equation gives $a(a^2 - 3b^2) = 2$. Further assuming that a and b may be integers, and realizing that the only factors of 2 are 1 and 2, Bombelli discovered that $a = 2$ and $b = 1$ solved the first equation. Since they also solve the second, he found that $(2 + \sqrt{-1})^3 = 2 + 11\sqrt{-1}$. Thus $x = (2 + \sqrt{-1}) - (-2 + \sqrt{-1}) = 4$.

By considering $\sqrt{-1}$ as an acceptable quantity, Bombelli found a real solution to the cubic equation. This legitimized complex numbers as a legitimate area of study.

John Wallis (English 1619 - 1703)
Attempts to view complex solutions to quadratic equations as points on a plane.

Abraham de Moivre (French 1667 - 1754)
Used complex numbers in his formula

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

Leonhard Euler (Swiss 1707 - 1783)
Understood DeMoivre's formula as giving solutions to the equation $x^n - 1 = 0$, viewed as vertices on a regular polygon.

Carl Friedrich Gauss (German 1777 - 1855)
Completed the geometric interpretation of the complex number $x+yi$ as the point (x, y) on the complex plane. Proved the Fundamental Theorem of Algebra.

Augustin-Louis Cauchy (French 1789 - 1857)
Formalized complex analysis and discovered many of its fascinating theorems.

1. Complex Algebra

Define addition and multiplication on the set $\mathbb{R}^2$ by

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

and

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

Let $\mathbb{C}$ denote the set $\mathbb{R}^2$ together with this addition and multiplication; we call $\mathbb{C}$ the set of *complex numbers*.

Let $f : \mathbb{R} \to \mathbb{C}$ be given by $f(x) = (x, 0)$. This embeds the real line into $\mathbb{C}$, in a manner which preserves addition and multiplication; we call the image the *real axis*, and identify $\mathbb{R}$ with its image.

Let $i = (0, 1)$. Then $i^2 = i \cdot i = (-1, 0) = -1$. We call $\{(0, y) \mid y \in \mathbb{R}\}$ the *imaginary axis*.

Every element of $\mathbb{C}$ can be written as $x + iy$ in a unique way, where $x, y \in \mathbb{R}$; that is,

$$\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}, i^2 = -1\}.$$

One can show that these operations have the following properties:

(F1) $a + b = b + a$ for every $a, b \in \mathbb{C}$;
(F2) $(a + b) + c = a + (b + c)$ for every $a, b, c \in \mathbb{C}$;
(F3) there exists $0 \in \mathbb{C}$ such that $a + 0 = a$ for every $a \in \mathbb{C}$;
(F4) for every $a \in \mathbb{C}$ there exists $b \in \mathbb{C}$ such that $a + b = 0$;
(F5) $ab = ba$ for every $a, b \in \mathbb{C}$;
(F6) $(ab)c = a(bc)$ for every $a, b, c \in \mathbb{C}$;
(F7) there exists $1 \in \mathbb{C}$ such that $a \cdot 1 = a$ for every $a \in \mathbb{C}$;
(F8) for every $a \in \mathbb{C} \setminus \{0\}$ there exists $c \in \mathbb{C}$ such that $ac = 1$;
(F9) $a(b + c) = ab + ac$ for every $a, b, c \in \mathbb{C}$.

Together, these properties state that $\mathbb{C}$ is a *field*. Note that

- $0 = 0 + i0$;
- $1 = 1 + i0$;
- $-(x + iy) = -x + i(-y) = -x - iy$;
- $(x + iy)^{-1} = \frac{x-iy}{x^2+y^2}$.

2. Complex Geometry

Let $z = x + iy$ be an arbitrary complex number. The *real part* of z is $\Re(z) = x$. The *imaginary part* of z is $\Im(z) = y$. We view $\mathbb{R}$ as the subset of $\mathbb{C}$ consisting of those elements whose imaginary part is zero.

We graph complex number on the xy-plane, using the real part as the first coordinate and the imaginary part as the second coordinate. Under this interpretation, the set $\mathbb{C}$ becomes a real vector space of dimension two, with scalar multiplication given by complex multiplication by a real number. We call this vector space the *complex plane*.

Thus the geometric interpretation of complex addition is vector addition.

Let $z = x + iy$ be an arbitrary complex number. The *conjugate* of z is $\overline{z} = x - iy$. This is the mirror image of z under reflection across the real axis. The *modulus* of z is $|z| = \sqrt{x^2 + y^2}$. This is the length of z as a vector. Note that $z\overline{z} = |z|^2$. The *angle* of z, denoted by $\angle(z)$, is the angle between the vectors $(1, 0)$ and (x, y) in the real plane $\mathbb{R}^2$; this is well-defined up to a multiple of 2π.

Let $r = |z|$ and $\theta = \angle(z)$. Then $x = r\cos\theta$ and $y = r\sin\theta$. Define a function

$$\text{cis} : \mathbb{R} \to \mathbb{C} \quad \text{by} \quad \text{cis}(\theta) = \cos\theta + i\sin\theta.$$

Then $z = r\text{cis}(\theta)$; this is the *polar representation* of z.

Recall the trigonometric formulae for the cosine and sine of the sum of angles:

$$\cos(A + B) = \cos A\cos B - \sin A\sin B$$

and

$$\sin(A + B) = \cos A\sin B + \sin A\cos B.$$

Let $z_1 = r_1\text{cis}(\theta_1)$ and $z_2 = r_2\text{cis}(\theta_2)$. Then

$$\begin{aligned} z_1z_2 &= r_1r_2(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= r_1r_2((\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)) \\ &= r_1r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)) \\ &= r_1r_2\text{cis}(\theta_1 + \theta_2). \end{aligned}$$

Thus the geometric interpretation of complex multiplication is:

(a) The radius of the product is the product of the radii;
(b) The angle of the product is the sum of the angles.

In particular, if $|z| = 1$, then $z = \text{cis}(\theta)$ for some θ, and $z^n = \text{cis}(n\theta)$. Restate this as

Theorem 1 (DeMoivre's Theorem)**.** $\text{cis}^n(\theta) = \cos(n\theta) + i\sin(n\theta)$.

Example 2. Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = 2z$. Then f dilates the complex plane by a factor of 2.

Example 3. Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = iz$. Then f rotates the complex plane by 90 degrees.

Example 4. Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = (1 + i)z$. Note that $|1 + i| = \sqrt{2}$ and $\angle(1 + i) = \frac{\pi}{4}$. Then f dilates the complex plane by a factors of $\sqrt{2}$ and rotates it by 45 degrees.

3. Complex Powers and Roots

Let $z = r\text{cis}(\theta)$ and let $n \in \mathbb{N}$. Then $z^n = r^n\text{cis}(n\theta)$.

The *unit circle* in the complex plane is

$$\mathbb{U} = \{z \in \mathbb{C} \mid |z| = 1\}.$$

Note that if $u_1, u_2 \in \mathbb{U}$, then $u_1 u_2 \in \mathbb{U}$.

Let $\zeta \in \mathbb{C}$ and suppose that $\zeta^n = 1$. We call ζ an n^{th} *root of unity.* If $\zeta^m \neq 1$ for $m \in \{1, \ldots, n-1\}$, we call ζ a *primitive* n^{th} *root of unity.*

Let $\zeta = \text{cis}(\frac{2\pi}{n})$. Then $\zeta^n = \text{cis}(n\frac{2\pi}{n}) = \text{cis}(2\pi) = 1$; one sees that ζ is a primitive n^{th} root of unity. Thus primitive roots of unity exist for every n. As m ranges from 0 to $n-1$, we obtain distinct complex numbers ζ^m, all of which are n^{th} roots of unity. These are all of the n^{th} roots of unity; thus for each $n \in \mathbb{N}$, there are *exactly* n distinct n^{th} roots of unity.

If one graphs the n^{th} roots of unity in the complex plane, the points lie on the unit circle and they are the vertices of a regular n-gon, with one vertex always at the point $1 = 1 + i0$.

Let $z = r\text{cis}(\theta)$. Then z has exactly n distinct n^{th} roots; they are

$$\sqrt[n]{z} = \sqrt[n]{r}\zeta_n^m\text{cis}(\frac{\theta}{n}), \quad \text{where } m \in \{0, \ldots, n-1\}.$$

The Fundamental Theorem of Algebra states that every polynomial with complex coefficients has a root in the complex numbers.

4. Complex Analysis

Let $f : \mathbb{C} \to \mathbb{C}$. We say that f is *continuous* at z_0 if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|z - z_0| < \delta \Rightarrow |f(z) - f(z_0)| < \epsilon$.

We say that f is *differentiable* at z_0 if the limit

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists.

Complex differentiability has some amazing consequences; for example, it can be shown that every complex differentiable function is analytic.

We use the Taylor series expansion for several real transcendental functions in order to define their complex counterparts.

Define the complex exponential function

$$\exp : \mathbb{C} \to \mathbb{C} \text{ by } \exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Define the complex sine function by

$$\sin : \mathbb{C} \to \mathbb{C} \text{ by } \sin(z) = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots$$

Define the complex cosine function by

$$\cos : \mathbb{C} \to \mathbb{C} \text{ by } \cos(z) = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots$$

Note that exp, sin, and cos, when restricted to $\mathbb{R} \subset \mathbb{C}$, are defined so as to be consistent with other definitions of these real functions.

Define $\log : \mathbb{C} \to \mathbb{C}$ to be an inverse function of exp. Let $w, z \in \mathbb{C}$. We define w^z by

$$w^z = \exp(z \log(w)).$$

Thus $\exp(z) = e^z$.

Euler evaluated $\exp(iz)$, separating the real and imaginary parts, and found

$$\begin{aligned}
\exp(iz) &= \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \\
&= 1 + iz + i^2\frac{z^2}{2!} + i^3\frac{z^3}{3!} + i^4\frac{z^4}{4!} + i^5\frac{z^5}{5!} + i^6\frac{z^6}{6!} + i^7\frac{z^7}{7!} + \dots \\
&= (1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots) + i(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots) \\
&= \cos z + i \sin z.
\end{aligned}$$

In particular, if $z = \theta \in \mathbb{R}$, we have

Theorem 5 (Euler's Theorem)**.** *Let $\theta \in \mathbb{R}$. Then*

$$e^{i\theta} = \text{cis}(\theta).$$

Letting $\theta = \pi$, we get the beautiful

$$e^{i\pi} + 1 = 0,$$

a formula that relates the four most important constants in mathematics.

5. Sum of Square Reciprocals

5.1. Historical Background. Recall the *triangular numbers*

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Leibnitz was challenged by Huygens to find the sum of their reciprocals. First factor out a 2 from all the terms $\frac{2}{n(n+1)}$; then compute

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n(n+1)} &= \sum_{n=1}^{\infty} \left[\frac{n+1}{n(n+1)} - \frac{n}{n(n+1)} \right] \\
&= \sum_{n=1}^{\infty} \left[\frac{1}{n} - \frac{1}{n+1} \right] \\
&= (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + (\frac{1}{3} - \frac{1}{4}) + (\frac{1}{4} - \frac{1}{5}) + \dots \\
&= 1 - (\frac{1}{2} - \frac{1}{2}) - (\frac{1}{3} - \frac{1}{3}) - (\frac{1}{4} - \frac{1}{4}) - \dots \\
&= 1.
\end{aligned}$$

Thus the sum of the reciprocals of the triangular numbers is 2.

Jacob Bernoulli, who knew that the harmonic series $\sum \frac{1}{n}$ diverges, then realized that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 1 + \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 2.$$

Euler was able to compute the value to which the sum of the reciprocals of the square natural numbers converges.

5.2. Polynomials with Specified Roots. Let $a_1, \dots, a_n \in \mathbb{C}$. We wish to construct a canonical polynomial with these zeros. One way is to select the polynomial to be *monic*; that is, to have 1 as the leading coefficient. The polynomial with this property is just

$$f(x) = \prod_{i=1}^{n} (x - a_i).$$

In this case, we know that the coefficients of $f(x)$ are symmetric functions of the zeros. However, we may also choose to normalize the polynomial by selecting the constant coefficient to be 1. For this case, set

$$(\dagger) \qquad g(x) = \prod_{i=1}^{n} (1 - \frac{x}{a_i}).$$

The coefficient of x in $g(x)$ is

$$(*) \qquad \sum_{i=1}^{n} \frac{-1}{a_i}.$$

5.3. Euler's Method. Let $g(x) = \frac{\sin x}{x}$; the power series expansion for $g(x)$ is arrived at by taking the Taylor series for $\sin x$ and dividing it, term by term, by x, to obtain:

$$g(x) = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \dots.$$

This has the appearance of a polynomial whose constant coefficient in 1, except that it infinitely many terms. Euler, being undeterred by this last fact, assumed that $g(x)$ could be written as an infinite product of linear terms as in equation (†).

Note that $g(0) = 1$; otherwise, the zeros of $g(x)$ are exactly those of $\sin x$; they are $Z = \{\pm\pi, \pm 2\pi, \pm 3\pi, \dots\}$. Thus Euler arrives at

$$\begin{aligned} g(x) &= \prod_{z \in Z} (1 - \frac{x}{z}) \\ &= \left((1 - \frac{x}{\pi})(1 + \frac{x}{\pi})\right)\left((1 - \frac{x}{2\pi})(1 + \frac{x}{2\pi})\right) \cdots \left((1 - \frac{x}{n\pi})(1 + \frac{x}{n\pi})\right) \cdots \\ &= \left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right) \cdots \left(1 - \frac{x^2}{n^2\pi^2}\right) \cdots \\ &= \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2\pi^2}\right). \end{aligned}$$

Multiplying out this infinite product, Euler finds the coefficient of the x^2 term, and equates it to the coefficient of the x^2 term of the power series expansion of $g(x)$, as in equation $(*)$, to get

$$-\frac{1}{3!} = \sum_{n=1}^{\infty} \frac{-1}{n^2\pi^2}.$$

Multiply both sides by $-\pi^2$ to arrive at the mysterious result

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

CHAPTER XIV

Fields

1. Fields

Definition 1. A *field* is a set F together with operations

$$+ : F \times F \to F \text{ and } \cdot : F \times F \to F$$

satisfying

(F1) $a + b = b + a$ for every $a, b \in F$;
(F2) $(a + b) + c = a + (b + c)$ for every $a, b, c \in F$;
(F3) there exists $0_F \in F$ such that $a + 0_F = a$ for every $a \in F$;
(F4) for every $a \in F$ there exists $b \in F$ such that $a + b = 0_F$;
(F5) $ab = ba$ for every $a, b \in F$;
(F6) $(ab)c = a(bc)$ for every $a, b, c \in F$;
(F7) there exists $1_F \in F$ such that $a \cdot 1_F = a$ for every $a \in F$;
(F8) for every $a \in F \setminus \{0_F\}$ there exists $c \in F$ such that $ac = 1_F$;
(F9) $a(b + c) = ab + ac$ for every $a, b, c \in F$;

Definition 2. Let F be a field. A *subfield* of F is a subset $S \subset F$ such that
(S0) $1 \in S$;
(S1) $a, b \in S \Rightarrow a + b \in S$;
(S2) $a \in S \Rightarrow -a \in S$;
(S3) $a, b \in S \Rightarrow ab \in S$;
(S4) $a \in S \Rightarrow a^{-1} \in S$.

If S is a subfield of F, we write $S \leq F$.

Remark 3. Properties **(S0)** through **(S4)** imply that a subfield of F is a subset of F which is itself a field.

Problem 1. Let F be a field and $\mathcal{S}$ be a collection of subfields of F.
Show that $\cap\mathcal{S} \leq F$.

Definition 4. Let $A \subset F$. The *subfield of F generated by A*is the intersection of all subfields of F which contain A.

If S is a subfield of F and $A \subset F$. let $S(A)$ denote the subfield of F generated by $S \cup A$. If $A = \{\alpha_1, \ldots, \alpha_n\}$ is finite, let $S(\alpha_1, \ldots, \alpha_n) = S(A)$. In particular, if $a \in F$, let $S(a) = S(\{a\})$.

Remark 5. Every subfield of $\mathbb{C}$ contains $\mathbb{Q}$, so every subfield generated by a subset of $\mathbb{C}$ contains $\mathbb{Q}$.

Example 6. Let $\alpha = \sqrt{2}$. Then

$$\mathbb{Q}(\alpha) = \{a + b\alpha \mid a, b \in \mathbb{Q}\}.$$

2. Polynomials

Definition 7. Let F be a field. A *polynomial over* F is a function $f : F \to F$ of the form

$$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n,$$

where n is a nonnegative integer and $a_i \in F$ for $i = 1, \ldots, n$, with $a_n \neq 0$ (unless $f(X) = 0$). We call the variable X an *indeterminate*.

The number n is called the *degree* of f, and is denoted by $\deg(f)$, The elements a_i are called the *coefficients* of f.

The number a_n is called the *leading coefficient*. We say that f is *monic* if $a_n = 1$.

The element a_0 is called the *constant coefficient*. The polynomials of degree zero are called *constants*, and are identified with the elements of the field F. By convention, $\deg(0) = -\infty$.

The set $F[X]$ is closed under addition, subtraction, and multiplication.

Proposition 8 (Division Algorithm for Polynomials)**.** *Let F be a field and let $f, g \in F[X]$. Then there exist polynomials $q, r \in \mathbb{F}[X]$ such that*

$$g = qf + r \quad \text{such that} \quad \deg(r) < \deg(f).$$

If f and g are monic, then q and r may be chosen to be monic or zero.

Proof. Without loss of generality, assume that f and g are monic. Let

$$S = \{h \in F[X] \mid h = g - qf \text{ for some monic } q \in F[X]\}.$$

Clearly S is nonempty; let $r \in S$ be a polynomial of minimal degree in S, so that $r = g - qf$ for some monic $q \in F[X]$. Then $g = qf + r$.

We claim that $\deg(r) < \deg(f)$, To see this, let $k = \deg(r) - \deg(f)$, and assume that $k \geq 0$. Then $X^k \in F[X]$, and $h = r - X^k f = g - (q - X^k)f \in S$ is a monic polynomial of degree less than that of r, contradicting the selection of r. □

Definition 9. Let F be a field and let $f, g \in F[X]$ We say that g is *divisible* by f, or that f is a *factor* of g, or that f *divides* g, and write $f \mid g$, if there exists $k \in F[X]$ such that $g = fk$. We see that f divides g if and only if the remainder upon division of g by f is $r = 0$.

Definition 10. Let F be a field, $f \in F[X]$, and $\alpha \in F$. If $\alpha \in F$, we say that α is a *zero* of f if $f(\alpha) = 0$. In this case, we say that f *annihilates* α.

Proposition 11 (Remainder Theorem)**.** *Let F be a field, $f \in F[X]$, and $\alpha \in F$. Let $h(X) = (X - \alpha) \in F[X]$. Write $f = hq + r$, where $\deg(r) < \deg(h)$. Then $r \in F$, and $f(\alpha) = r$.*

Proposition 12 (Factor Theorem)**.** *Let F be a field, $f \in F[X]$, and $\alpha \in F$. Let $h(X) = (X - \alpha) \in F[X]$. Then $h \mid f$ if and only if $f(\alpha) = 0$.*

Proposition 13. *Let F be a field and let $\alpha \in F$. Suppose that $g = fq$ for some $f, g, q \in F[X]$, and that $g(\alpha) = 0$. Then either $f(\alpha) = 0$ or $q(\alpha) = 0$.*

Definition 14. Let $f, g \in F[X]$. A *greatest common divisor* of m and n, denoted $\gcd(m, n)$, is a monic $d \in F[X]$ such that

(a) $d \mid f$ and $d \mid g$;
(b) If $e \mid f$ and $e \mid g$, then $e \mid d$.

Proposition 15 (Euclidean Algorithm for Polynomials)**.** *Let $f, g \in F[X]$. Then there exists $d \in F[X]$ such that $d = \gcd(m, n)$, and there exist $s, t \in F[X]$ such that*

$$d = sf + tg.$$

If f and g are monic, we may choose s and t to be monic.

Proof. Without loss of generality, assume that f and g are monic. Let

$$S = \{h \in F[X] \mid h = sf + tg \text{ for some monic } s, t \in F[X]\}.$$

Clearly S is nonempty; select $d \in S$ of minimal degree, so that $d = sf + tg$ for some monic $s, t \in F[X]$.

Now $f = qd + r$ for some monic $q, r \in F[X]$ with $\deg(r) < \deg(d)$. Then $f = q(sf + tg) + r$, so $r = (1 - qs)f + (qt)g \in S$. If r is nonzero, this contradicts the selection of d; thus $r = 0$, which shows that $d \mid f$. Similarly, $d \mid g$.

If $e \mid f$ and $e \mid g$, then $f = ke$ and $g = le$ for some $k, l \in F[X]$. Then $d = ske + tle = (sk + tl)e$. Therefore $e \mid d$. This shows that $d = \gcd(m, n)$. □

Definition 16. Let F be a field and let $f \in F[X]$. We say that f is *irreducible over F* if whenever $f = gh$ for some $g, h \in F[X]$, either $\deg(g) = 1$ or $\deg(h) = 1$.

Example 17. If $\deg(f) \in \{2, 3\}$, then f is irreducible over F if and only if f has no zero in F.

3. Field Extensions

Definition 18. A *field extension* E/F consists of a field E which contains a field F.

Definition 19. Let E/F be a field extension, and let $\alpha \in E$. We say that α is *algebraic* over F if there exists a nonzero polynomial $f \in F[X]$ such that $f(\alpha) = 0$. Otherwise, we say that α is *transcendental* over F.

Proposition 20. *Let E/F be a field extension and let $\alpha \in E$ be algebraic over F. Then there exists a unique monic irreducible polynomial $f \in F[X]$ such that $f(\alpha) = 0$.*

Proof. Since α is algebraic over F, there exists some polynomial in $F[X]$ which annihilates α. Let $f \in F[X]$ be a nonzero polynomial of minimal degree which annihilates α. Clearly f is irreducible, since it is of minimal degree. We may divide by the leading coefficient to see that we may select f to be monic. Now suppose that g is another monic polynomial of minimal degree which annihilates α. We have $\deg(f) = \deg(g)$. Then $\deg(f - g) < \deg(f) = \deg(g)$. Since f is of minimal degree among nonzero polynomials which annihilate α, we must have $f - g = 0$. Thus $f = g$, and f is unique. □

Definition 21. Let E/F be a field extension and let $\alpha \in E$ be algebraic over F. The *minimum polynomial* of α over F, denoted $\text{minpoly}(\alpha/F)$, is the unique monic irreducible polynomial which annihilates α. The *degree* of α over F, denoted $\deg(\alpha/F)$, is equal to $\deg(\text{minpoly}(\alpha/F))$.

Definition 22. Let E/F be a field extension and let $\alpha \in E$. The *evaluation map* on $F[X]$ with respect to α is the function $\psi_\alpha : F[X] \to E$ defined by $f \mapsto f(\alpha)$. The image of the evaluation map is denoted $F[\alpha]$; that is,

$$F[\alpha] = \psi_\alpha(F[X]) = \left\{ \sum_{i=0}^{k} a_i \alpha^i \mid k \in \mathbb{N}, a_i \in F \right\} \subset E.$$

Proposition 23. *Let E/F be a field extension and let $\alpha \in E$. If α is transcendental over F if and only if ψ_α is injective.*

Proof. Suppose that α is transcendental. Let $f, g \in F[X]$ so that $f(\alpha)$ and $g(\alpha)$ are arbitrary members of $F[\alpha]$. Suppose that $f(\alpha) = g(\alpha)$; then $(f-g)(\alpha) = 0$, so $(f-g)$ is a polynomial which annihilates α. Since α is transcendental, we must have $f - g = 0$, so $f = g$.

On the other hand, if α is not transcendental, it is algebraic; let $f = \text{minpoly}(\alpha/F)$. Then $\psi_\alpha(f) = \psi_\alpha(0)$, and ψ_α is not injective. □

Proposition 24. *Let E/F be a field extension and let $\alpha \in E$. Let $F[\alpha] = \psi_\alpha(F[X])$ denote the image of $F[X]$ under the evaluation map. Let α is algebraic over F and $\deg(\alpha/F) = n$, then $F[\alpha] = S$, where*

$$S = \left\{ \sum_{i=0}^{n-1} a_i \alpha^i \mid a_i \in F \right\};$$

moreover, $F[\alpha]$ is a field, and $F[\alpha] = F(\alpha)$.

Proof. Clearly all elements of the form $\sum_{i=0}^{n-1} a_i \alpha^i$ are in $F[\alpha]$, so $S \subset F[\alpha]$.

Let $f \in F[X]$ be the minimum polynomial of α over F. Let $g \in F[X]$; then $g(\alpha)$ is an arbitrary member of $F[\alpha]$. Now $g(X) = f(X)q(X) + r(X)$, where $\deg(r) < \deg(f)$. By the remainder theorem, $g(\alpha) = f(\alpha)q(\alpha) + r(\alpha) = r(\alpha) \in S$.

Since $F[X]$ is closed under addition, subtraction, and multiplication, so is $F[\alpha]$. We only need to show that $f(\alpha)$ if invertible for $f(\alpha) \neq 0$.

Let $\beta \in F[\alpha]$. Then $\beta = g(\alpha)$ for some $g \in F[X]$; by the division algorithm, we may select g so that $\deg(g) < \deg(f)$. Since f is irreducible, we see that $\gcd(f, g) = 1$, so there exist $s, t \in F[X]$ such that $sf + tg = 1$. Then $t(\alpha)g(\alpha) = 1$, so $\beta^{-1} = t(\alpha)$, and β is invertible. □

4. Vector Spaces

Definition 25. Let F be a field. A *vector space* over F is a set V together with operations

$$+ : V \times V \to V \quad \text{and} \quad \cdot : F \times V \to V$$

satisfying

(V1) $v + w = w + v$ for all $v, w \in V$;
(V2) $v + (w + x) = (v + w) + x$ for all $v, w, x \in V$;
(V3) there exists $0_V \in V$ such that $v + 0_V = v$ for all $v \in V$;
(V4) for every $v \in V$ there exists $w \in V$ such that $v + w = 0_V$;
(V5) $1_F \cdot v = v$ for every $v \in V$;
(V6) $(ab)v = a(bv)$ for every $v \in V$ and $a, b \in F$;
(V7) $(a + b)v = av + bv$ for every $v \in V$ and $a, b \in F$.
(V8) $a(v + w) = av + aw$ for every $v, w \in V$ and $a \in F$;

Problem 2. Let V be a vector space over a field F. Let $a \in F$ and $x \in V$.
(a) Show that $0_F \cdot x = 0_V$.
(b) Show that $a \cdot 0_V = 0_V$.
(c) Show that $(-1_F) \cdot x = -x$.

Definition 26. Let V be a vector space over a field F.

A *subspace* of V is a subset $W \subset V$ such that

(W0) $0_V \in W$;
(W1) $x, y \in W \Rightarrow x + y \in W$;
(W2) $a \in F, x \in W \Rightarrow ax \in W$.

If W is a subspace of V, this is denoted by $W \leq V$.

Remark 27. Properties **(W0)** through **(W2)** imply that a subspace of V is a subset of V which is itself a vector space.

Problem 3. Let V be a vector space over a field F and let $\mathcal{W}$ be a collection of subspaces of V.
Show that $\cap\mathcal{W} \leq V$.

Definition 28. Let V be a vector space over a field F and let $A \subset V$. The *subspace of V generated by A*, denoted $\mathrm{gv}_V(A)$, the intersection of all subspaces of V which contain A. This subspace is called the *span* of A.

Problem 4. Let V be a vector space over a field F and let $A = \{v_1, \ldots, v_n\}$. Show that

$$\mathrm{gv}_V(A) = \left\{ \sum_{i=1}^{n} a_i v_i \mid a_i \in F \right\}.$$

5. Vector Space Dimension

Definition 29. Let V be a vector space over a field F. Let $B \subset V$.

We say that B *spans* V is for every $x \in V$ there exist $a_1, \ldots, a_n \in F$ and $v_1, \ldots, v_n \in B$ such that $x = \sum_{i=1}^{n} a_i v_i$.

We say that B is *linearly independent* if whenever $v_1, \ldots, v_n \in B$ are distinct elements of B and $a_1, \ldots, a_n \in F$,

$$\sum_{i=1}^{n} a_i v_i = 0 \Rightarrow a_i = 0 \text{ for } i = 1, \ldots, n.$$

We say that B is a *basis* for V if B spans V and is linearly independent.

Problem 5. Let V be a vector space over a field F and let $X \subset V$ span V. Show that $V = \mathrm{gv}_V(X)$.

Problem 6. Let V be a vector space over a field F and let $X \subset V$ be linearly independent. Let $v \in X$. Show that $\mathrm{gv}_V(X \setminus \{v\})$ is a proper subset of $\mathrm{gv}_V(X)$.

Problem 7. Let V be a vector space over a field F and let $X \subset V$ span V. Show that there exists a subset $B \subset X$ such that B is a basis for V.

Problem 8. Let V be a vector space over a field F and let $X \subset V$ be linearly independent. Show that there exists a subset $Y \subset V$ such that $X \cup Y$ is a basis for V.

Problem 9. Let V be a vector space over a field F. Let $A = \{v_1, \ldots, v_m\}$ and $B = \{w_1, \ldots, w_n\}$ be bases for V. Show that $m = n$.

Definition 30. Let V be a vector space over a field F. If V has a basis containing n elements, where $n \in \mathbb{N}$, we say that V is *finite dimensional*, and that n is the *dimension* of V; this is denoted by $\dim(V) = n$.

Problem 10. Let V be a vector space over a field F and let $U, W \leq V$. Set $U + W = \{u + w \mid u \in U, w \in W\}$.
(a) Show that $U + W \leq V$.
(b) Show that $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$.

Problem 11. Let F be a field and let n be a positive integer. Let F^n denote the cartesian product of F with itself n times. Show that F^n is a vector space over F of dimension n.

Observation 31. *Let E/F be a field extension. We may add the elements of E, and multiply them by elements of F. In this way, we may view E as a vector space over F.*

Definition 32. Let E/F be a field extension. The *degree* of the extension, denoted $[E : F]$, is its dimension of E as a vector space over F.

6. Types of Extensions

Definition 33. Let E/F be a field extension.

We say that E/F is a *primitive extension* if $E = F[\alpha]$ for some $\alpha \in E$ which is algebraic over F.

We say that E/F is a *finite extension* if $[E : F] < \infty$.

We say that E/F is a *algebraic extension* if every element of E is algebraic over F.

Proposition 34. *Let E/F be a primitive extension such that $E = \mathbb{F}[\alpha]$, where α is algebraic over F with* $\text{minpoly}(\alpha/F) = f \in F[X]$. *Let $n = \deg(f)$. Then the set*

$$B = \{1, \alpha, \ldots, \alpha^{n-1}\}$$

is a basis for E/F, and in particular, $[E : F] = n$.

Proof. Since $E = F[\alpha]$, that B spans E is a direct consequence of Proposition 24. To see that B is linearly independent, let

$$a_0 \cdot 1 + a_1\alpha + \cdots + a_n\alpha^{n-1} = 0$$

be a dependence relation. Then α is a root of the polynomial $\sum_{i=1}^{n-1} a_i X^i$. Since this polynomial has lower degree than f, it must be the zero polynomial, so $a_i = 0$ for every i. This shows that B is linearly independent over F. □

Proposition 35. *Let E/F be a finite extension. Then E/F is an algebraic extension.*

Proof. Let $[E : F] = n$, and let $\alpha \in E$. The set $S = \{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ contains $n + 1$ elements, and so it must be linearly dependent over F. Thus there exists a nontrivial dependence relation

$$a_0 \cdot 1 + a_1\alpha + \cdots + a_n\alpha^n = 0.$$

Let $f(X) = a_0 + a_1X + \ldots a_nX^n$. Then $f(\alpha) = 0$, so α is algebraic over F. □

Proposition 36. *Let K/E and E/F be finite field extensions of dimension n and m respectively. If $\{z_1, \ldots, z_n\}$ is a basis for K/E and $\{y_1, \ldots, y_m\}$ is a basis for K/F, then $\{y_iz_j \mid i = 1, \ldots, m;\ j = 1, \ldots, n\}$ is a basis for K/F. In particular, K/F is finite, and*

$$[K : F] = [K : E][E : F].$$

Proof. Let $\alpha \in K$. Then α is in the span of $\{z_j\}$, so $\alpha = \sum_{j=1}^{n} b_jz_j$ for some $b_j \in E$. Since each $b_j \in E$, it is in the span of $\{y_i\}$, so $b_j = \sum_{i=1}^{m} a_{ij}y_i$ for some $a_{ij} \in F$. Thus

$$\alpha = \sum_{j=1}^{n} \left[\sum_{i=1}^{m} a_{ij}y_i \right] z_j = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij}y_iz_j.$$

Thus $\{y_iz_j\}$ spans K.

Now consider a dependence relation $\sum_{j=1}^{n} \sum_i i = 1^m a_ijy_iz_j = 0$. Collect like terms to obtain $\sum_{j=1}^{n} \left[\sum_{i=1}^{m} a_{ij}y_i \right] z_j = 0$. Since $\{z_j\}$ is linearly independent, we must have $\sum_{i=1}^{n} a_{ij}y_i = 0$ for every j. But since $\{y_i\}$ is linearly independent, this implies that $a_{ij} = 0$ for every i and j. Thus $\{y_iz_j\}$ is linearly independent over F. □

7. Field of Constructible Numbers

Definition 37. Let $S \subset \mathbb{C}$ and set $z \in \mathbb{C}$. We say that a line $L \subset \mathbb{C}$ is constructible from S if $L \cap S$ contains at least two points. We say that a circle $C \subset \mathbb{C}$ is constructible from S if the center of C is in S and $C \cap S$ is nonempty. We say a point $z \in \mathbb{C}$ is constructible from S if one of the following conditions holds:

- **(C0)** $z \in S$;
- **(C1)** $z \in L_1 \cap L_2$, where L_1 and L_2 are lines constructible from S;
- **(C2)** $z \in L_1 \cap C_1$, where L_1 is a line and C_1 is a circle constructible from S;
- **(C3)** $z \in C_1 \cap C_2$, where C_1 and C_2 are circles constructible from S.

Let $C(S)$ be the set of points which are constructible from S.

Set $C_0(S) = S$ and inductively set $C_{n+1}(S) = C(C_n(S))$. Let $S = \{0, 1\} \in \mathbb{C}$, and define

$$\mathbb{K} = \cup_{n=0}^{\infty} C_n(S);$$

members of $\mathbb{K}$ are called *constructible numbers.*

Proposition 38. *Let $a, b \in \mathbb{K}$. Then*

- **(K1)** $a + b \in \mathbb{K}$;
- **(K2)** $-a \in \mathbb{K}$;
- **(K3)** $ab \in \mathbb{K}$;
- **(K4)** $a^{-1} \in \mathbb{K}$ *if* $a \neq 0$;
- **(K5)** $\pm\sqrt{a} \in \mathbb{K}$;
- **(K6)** $\overline{a} \in \mathbb{K}$;

Thus the set $\mathbb{K}$ is a subfield of $\mathbb{C}$ which is closed under square roots and conjugation.

Proof. Note that $a+b$ is the fourth point in a parallelogram with points a, 0, and b; we have seen that this construction is possible. Also, $-a$ is the intersection of the line through 0 and a with the circle centered at 0 through a, so $-a$ is constructible.

Let $a = re^{i\theta}$ be the polar expression of a. Now $r = |a|$; this may be constructed by intersecting the real axis with the circle centered at 0 through a.

Now let $a = re^{i\theta}$ and $b = se^{i\gamma}$; then $ab = rse^{i(\theta+\gamma)}$. We have seen that if we can construct lengths r and s, then we can construct the length rs. We only need to show that we can construct the angle $\theta + \gamma$. Try to do this geometrically; otherwise it will follow algebraically from the similar facts for the real and imaginary parts of a and b.

Next we describe how to construct the conjugate $\overline{a}$ of a. Form the line perpendicular to the real axis and passing through a. Intersect this line with the circle centered at 0 through a. One point of intersection is a, the other is $\overline{a}$.

Consider that $a^{-1} = \frac{1}{r}e^{-i\theta}$. We have seen that we can construct $\frac{1}{r}$, and we can bisect any angle. Thus $a^{-1} \in \mathbb{K}$. □

Proposition 39. *Let $z \in \mathbb{C}$. Then $z \in \mathbb{K}$ if and only if $\Re z \in \mathbb{K}$ and $\Im z \in \mathbb{K}$. In particular, i is constructible.*

Proof. Note that the real axis is immediately constructible from $\{0, 1\}$, and the imaginary axis is constructible as the perpendicular to the real axis through 0.

Suppose that $z \in \mathbb{K}$. Then $|z|$ is the positive real number obtained as the intersection of real line and the circle centered at 0 and through z. Then $|z|^2$ is constructible since $\mathbb{K}$ is a field, and since $z\overline{z} = |z|^2$, we see that $\overline{z} = \frac{|z|^2}{z}$ is constructible. Thus $\Re z = \frac{1}{2}(z + \overline{z})$ is constructible, and $\Im z = z - \Re z$ is constructible.

Suppose that $\Re z$ and $\Im z$ are constructible. Now i is the intersection of the unit circle and the imaginary axis, so i is constructible. Thus $z = \Re z + i\Im z$ is constructible. □

8. Constructed Fields

Definition 40. Let $\boldsymbol{z} = (z_1, \ldots, z_n)$ be an n-tuple of complex numbers. We say that $\boldsymbol{z}$ is *constructed* if $z_1 = i$ and $z_{i+1} \in C(\mathbb{Q}[z_1, \ldots, z_i])$ for $i = 1, \ldots, n$. If $F \leq \mathbb{C}$, we say that F is constructed if $F = \mathbb{C}[z_1, \ldots, z_n]$ for some constructed tuple $(z_1, \ldots, z_n)$.

Proposition 41. *Let $F \leq \mathbb{C}$ and $z \in \mathbb{C}$. Suppose $i \in F$. Then $z \in F$ if and only if $\Re z, \Im z \in F$. In this case, $\overline{z} \in F$ and $|z|^2 \in F$.*

Proof. Let $z = x + iy$, where $x, y \in \mathbb{R}$. If $x, y, i \in F$, then obviously $z \in F$.

Suppose $z, i \in F$; then $z - iz \in F$. Now $z - iz = (x - ix) - (y - iy) = (x - y)(1-i)$. Since $i \in F$, $1-i \in F$, so $x-y \in F$. Now $(x-y)-z = y-iy = y(1-i)$, so $y \in F$. Thus $x \in F$. Now $\overline{z} = x - iy \in F$, so $|z|^2 = z\overline{z} \in F$. □

Proposition 42. *If $\alpha \in \mathbb{K}$, then there exists a constructed tuple $(z_1, \ldots, z_n)$ such that $\alpha = z_n$.*

Proof. It follows from the definition of constructibility that α can be constructed from finitely many stages from the set $\{0, 1\} \subset \mathbb{Q}$. The result follows from this. □

Proposition 43. *Let E/F be a field extension with $[E : F] = n$, and let $\alpha \in E$. Then $\deg(\alpha/F)$ divides n.*

Proof. We know that $[F[\alpha] : F] = \deg(\alpha/F) = \deg(\text{minpoly}(\alpha/F))$. By the product of degrees formula, $[E : F] = [E : F[\alpha]] \cdot [F[\alpha] : F]$. The result follows. □

Proposition 44. *Let E be a constructed field. Then $[E : \mathbb{Q}]$ is a power of two.*

Proof. Since E is a constructed field, there exists a constructed tuple $(z_1, \ldots, z_n)$, with $z_1 = i$, such that $E = \mathbb{Q}[z_1, \ldots, z_n]$, with $z_{i+1} \in \mathbb{Q}[z_1, \ldots, z_i]$.

Let $F_i = \mathbb{Q}[z_1, \ldots, z_i]$ for $i = 1, \ldots, n$; note $E = F_n$. We proceed by induction on n.

For $n = 1$, we have $z_1 = i$. Now $\text{minpoly}(i/\mathbb{Q}) = X^2 + 1$, and $\deg(z_1/\mathbb{Q}) = 2$, so the proposition is true in this case.

Now suppose that $n > 1$, and let $F = F_{n-1}$ and $\alpha = z_n$. By induction, $[F : \mathbb{Q}]$ is a power of two. We also know that $i \in F$, so $z \in F$ if and only if $\Re z, \Im z, \overline{z}, |z|^2 \in F$.

Since α is constructible from F, it is the intersection of lines and circles given by points in F.

Case 1: α is the point of intersection of two lines given by F.
Note that the slope of a line through two points in F is also in F; let $y = m_1 x + b_1$ and $y = m_2 x + b_2$ be lines which intersect at α, where $m_1, b_1, m_2, b_2 \in F$. Then the point of intersection is the complex number $\alpha = \frac{b_2 - b_1}{m_1 - m_2} + \frac{m_1 b_2 - b_1 m_2}{m_1 - m_2} i$, whose real and imaginary parts are in F, so $\alpha \in F$ in this case, and $\deg(\alpha/F) = 1$.

Case 2: α is a point of intersection of a line and a circle given by F.
Let $y = mx + b$ and $(x - h)^2 + (y - k)^2 = r^2$ be the equations of the line and the circle. Now $m, b \in F$. Since $w = h + ki$ is the center of the circle, $h, k \in F$. Also there exists a point $z \in \mathbb{C}$ whose distance from w is r, so $r = |w - z| \in F$. Substitution gives $(x - h)^2 + (mx + b - k)^2 - r^2 = 0$; this is a quadratic equation whose solution is of the form $x = A + B\sqrt{D}$, where $A, B, D \in F$. Let $y = mx + b$; now $\alpha = x + yi$, and since $x, y \in F[\sqrt{D}]$, so is α.

Case 3: α is a point of intersection of two circles given by F.
Subtracting the equations of the circles cancels both the x^2 and the y^2 terms, producing a linear equation in x and y. Use this in combination with the equation of one of the circles to reduce to Case 2. □

Proposition 45. *Let $\alpha \in \mathbb{C}$ be constructible. Then there exist $p \in \mathbb{N}$ such that* $\deg(\text{minpoly}(\alpha/\mathbb{Q})) = 2^p$.

Proof. If α is constructible, there exists a constructed tuple $(z_1, \ldots, z_n)$ such that $\alpha = z_n$. Let $E = \mathbb{Q}[z_1, \ldots, z_n]$; then $\alpha \in E$ and $[E : F]$ is a power of two. By a previous proposition, $\deg(\alpha/\mathbb{Q})$ divides $[E : F]$, so it is also a power of two. □

Proposition 46. *It is impossible to double a cube.*

Proof. Start with a cube whose sides have length one. To construct a cube with double the volume, one must be able to construct an edge of this cube; this requires the constructibility of the number $\alpha = \sqrt[3]{2}$.

The minimum polynomial of α over $\mathbb{Q}$ is $X^3 - 2$, so $\deg(\alpha/\mathbb{Q}) = 3$. Since 3 is not a power of 2, α is not constructible. □

APPENDIX A

Archimedes and the Canned Sphere

1. Introduction

1.0.1. *Historical Background.* Mathematics always precedes physics. The mathematics may be developed centuries before it is used, as is the case with Apollonius (ca. 240 B.C.) and his abstract treatise *Conic Sections*, which preceded the usage by Copernicus and Kepler of ellipses to explain the orbits of planets. Or the mathematics may be created by the physicist in order to solve his physical problems, as was the case with Newton. Either way, the physics cannot go forward until the mathematical model comes into existence.

Physics precedes engineering. What is physically possible is understood before what is physically practical can be implemented.

Thus one wonders what the world would be like if we kicked forward the pace of mathematical research during its reinvigoration in the fifteenth and sixteenth century, say by 200 years. Would the technology necessary to prevent famine have kept ahead of the growth of population? Would the great world wars have been prevented?

Between the ancient Greeks and the early Renaissance, European civilization was dominated first by the Roman Empire and then by the Holy Roman Church. Science as we understand it dissolved during this period, and much that was known to the Greeks was lost. After this long drought, during the 1400's, many brilliant people were being allowed to think and communicate their thoughts again, and a period of reconstruction began.

The details of calculus, necessary for the advancement of physics, were worked out during the late 1600's; thus it took nearly 200 years to advance to this point.

1.0.2. *Discovery of the Palimpsest.* In the first decade of the twentieth century, a Danish philologist Johan Ludvig Heiberg (1854-1928) discovered an ancient prayer book in a library in Constantinople. Barely visible behind the Latin text were Greek symbols. The book was an ancient parchment palimpsest.

Parchment is a material for the pages of a book, made from fine calf skin, sheep skin or goat skin. A palimpsest is a manuscript page, scroll, or book that has been written on, scraped off, and used again.

Heiberg came to realize that the Greek writing was a previously unknown work of the ancient Greek genius Archimedes, who lived in the third century BC, was written in the 10th century. In the 12th century it was imperfectly erased in order that a liturgical text could be written on the parchment, and Archimedes' work is still legible today. It was a book of nearly 90 pages before being made a palimpsest of 177 pages; the older leaves were folded so that each became two leaves of the liturgical book.

Heiberg was not allowed to take the book from the library, so he had every page photographed. Using only a magnifying glass, he attempted to read the Greek text, and published what he could. Shortly thereafter it was translated into English by Thomas Heath.

1.0.3. *Contents of the Palimpsest.* This book is known as *The Method*, or *Method Concerning Mechanical Theorems*, and it describes the process by which Archimedes discovered many of his results.

Although the only mathematical tools at its author's disposal were what we might now consider secondary-school geometry, he used those methods with rare brilliance, explicitly using infinitesimals to solve problems that would now be treated by integral calculus. Among those problems were that of the center of gravity of a solid hemisphere, that of the center of gravity of a frustum of a circular paraboloid, and that of the area of a region bounded by a parabola and one of its secant lines. Contrary to historically ignorant statements found in some 20th-century calculus textbooks, he did not use anything like Riemann sums, either in the work embodied in this palimpsest or in any of his other works. For explicit details of the method used, see how Archimedes used infinitesimals.

Historian Reviel Netz of Stanford University, with technical assistance from several persons at the Rochester Institute of Technology, has been trying to fill in gaps in Heiberg's account. In Heiberg's time, much attention was paid to Archimedes' brilliant use of infinitesimals to solve problems about areas, volumes, and centers of gravity. Less attention was given to the Stomachion, a problem treated in the Palimpsest that appears to deal with a children's puzzle. Netz has shown that Archimedes found that the number of ways to solve the puzzle is 17,152. This is perhaps the most sophisticated work in the field of combinatorics in classical antiquity.

2. Archimedes Biography

287 B.C. Archimedes was born in Sicily around 287 B.C. His father was an astronomer and mathematician named Phidias. Unfortunately, little else is known about Archimedes' early life. It is believed, however, that Archimedes' family was a rich and noble one, perhaps related in some way to Hiero, King of Syracuse.

269 B.C. Archimedes travelled to Egypt to study at Alexandria. This city had been founded by Alexander the Great in 331 B.C, and by 300 BC was home to 500,000 people. Alexandria was also the home of Euclid, who lived from about 330 to 275 B.C. Euclid was a renowned mathematician and may best be remembered for his book, "The Elements" which was the most important geometry book in the world for over 2000 years. Archimedes undoubtedly studied this book along with others in the great library of Alexandria, which contained more than a million books in the form of scrolls of papyrus.

263 B.C. Archimedes returned to Syracuse after his studies in Alexandria and settled down to a life of study and research. He would typically sit for hours pondering geometry diagrams drawn in the sand floor of his home or on papyrus scrolls. His experimentations soon made him indispensable to King Hiero, and ultimately, to the rest of the world.

Archimedes' abilities were put to good use by King Hiero. In one case, the hold of a huge boat made for the King had become full of water after a heavy rain.

Not sure how to remove the water from the ship, King Hiero asked Archimedes for assistance. Archimedes created what is now know as the "Archimedes Screw". It is a machine consisting of a hollow tube containing a spiral that could be turned by a handle at one end. When the lower end of the tube was put into the hold and the handle turned, water was carried up the tube and over the side of the ship. The "Archimedes Screw" soon became popular in Egypt as a device for irrigating fields and in other forms, is still in use today.

King Hiero had commissioned a new royal crown for which he provided solid gold to the goldsmith. But when the crown arrived, King Hiero was suspicious that the goldsmith only used some of the gold, kept the rest for himself and added silver to make the crown the correct weight. Archimedes was asked to determine whether or not the crown was pure gold without harming it in the process. Archimedes was perplexed but found inspiration while taking a bath. While noticing that the water overflowed from the tub when he lowered himself into it, he realized that he could measure the crown's density if he could determine the amount of water it displaced, or its "volume". Legend has it that Archimedes was so exuberant about his discovery that he ran down the streets of Syracuse naked shouting, "Eureka!" which meant "I've found it!" in Greek.

Archimedes found that the crown was indeed a fake proving that the goldsmith had cheated.

King Hiero relied on Archimedes' inventions for use in the military during a time when there was great competition for power in the Mediterranean region between Syracuse, Carthage and Rome. Putting his theories of levers and pulleys to work, Archimedes built other machines designed to defend Syracuse.

216 B.C. King Hiero died in the year 216 B.C. and was succeeded by his 15-year-old grandson Hieronymos. The new King formed an alliance with Hannibal, the ruler of Carthage, which alarmed the pro-Roman faction within Syracuse.

215 B.C. Hieronymos was assassinated in the Greek city of Leontini, ending his 13-month reign. After the assassination of Hieronymos, civil war erupted in Syracuse between the pro-Carthaginian and pro-Roman factions, during which most of Hiero's family was killed. The pro-Carthaginian faction was eventually victorious and two brothers of mixed Carthaginian-Syracusan descent, Hippokrates and Epikydes, took control of the city.

214 B.C. Marcellus led the Roman army in an invasion of Syracuse but they were thwarted by the ingenuity of Archimedes. Among his many inventions were the huge curved mirrors placed on top of the city walls. When the Roman fleet was in sight the mirrors were turned to reflect the Sun's rays onto the ships. The heat was so great that many ships burst into flames. Other ships were destroyed by huge boulders thrown by the catapults designed by Archimedes.

With the help of Archimedes' incredible machines, Syracuse was protected from the Roman army. One of these machines operated with great iron claws that could seize boats by the prow, draw them up into the air, and plunge them into the depths of the sea. Another projected huge wooden beams from the island's ramparts to gouge the hulls of enemy ships.

Unable to penetrate the devices which Archimedes had placed around the borders of Syracuse, Marcellus ultimately surrounded the city and prevented supplies from entering or leaving. The siege lasted over two years. Eventually,

in 212 B.C., the Romans took advantage of an unguarded section of the city walls and invaded the city.

212 B.C. During the siege of Syracuse, a Roman soldier burst through the door of Archimedes and demanded that the great military genius accompany him to the quarters of General Marcellus. Not realizing that the city had been invaded, Archimedes refused, claiming he had yet to finish a mathematical problem that presently occupied his attention. The soldier, in anger, struck the 75-year-old Archimedes dead.

Marcellus was distressed upon hearing the news of the death, and ordered that Archimedes be buried with honor. His tombstone was, as he wished, engraved with the geometrical diagram showing a sphere inside a cylinder, to remind the world of his great discoveries.

3. The Method's Journey

4th century A.D. During his life, Archimedes wrote out his theories on papyrus scrolls. Succeeding generations preserved his works by copying and recopying them onto other scrolls. Somewhere, in the fourth century A.D., scribes began to copy onto parchment, then bind them between wooden boards. This was the earliest version of what's known today as the "book".

10th century A.D. The Archimedes manuscript was copied onto parchment sheets and bound between wooden boards. Although manufactured more than a thousand years after the great mathematician's death, this book, which is now in the care of The Walters Art Gallery, is the earliest copy of Archimedes' treatises to survive.

12th century A.D. Parchment was scarce and it was common practice to re-use old manuscripts for newer writings. Apparently, the Archimedes text was taken apart, most likely in Constantinople, for this purpose. A scribe disassembled the manuscript and scraped off as much of the Archimedes text as he could. He cut the leaves in half along the inner fold and turned the page leaves 90 degrees before folding them in half. This scribe ruled fresh lines and copied new religious text onto the parchment, creating what's known as a "Palimpsest", or a text on parchment which has been overwritten with other text.

12th - 19th century A.D. Once the manuscript had become a religious text, it was considered a sacred document and cared for in the Holy Land, between Jerusalem and the Dead Sea. One of its homes was the monastery of Mar Saba, historically an intellectual and spiritual center for the Greek Church. The book was most likely used as religious text by the monastery's inhabitants for at least 400 years.

Early 1800's The palimpsest was moved from the monastery to the library of the Greek Patriarch in the Christian quarter of old Jerusalem. The book did not remain there long, however, as it continued to travel in the highest of religious circles. It is believed that the book travelled to the Church of the Holy Sepulchre, because it ultimately ended up in the Church's daughter house, the Metochion in Constantinople the city where the manuscript had first been created.

1846 A.D, Biblical scholar Constantine Tischendorf visited the Metochion Of The Holy Sepulchre to study the library's substantial collection of manuscripts. At the time, he claimed to find nothing of particular interest,

except for a palimpsest dealing with mathematics. Though he didn't quite understand the importance of his discovery, he must have sensed the book's value, because he acquired one of its leaves now owned by the Cambridge University Library in England.

1907 A.D. Danish philologist Johan Heiburg meticulously transcribed the manuscript using nothing but a magnifying glass. It's not known whether Heiberg suspected the palimpsest's true origins at first, but he ultimately realized that this ancient manuscript was indeed a previously unknown treatise by Archimedes, the great mathematician. His great achievement and extraordinary find made headlines in the New York Times on July 16, 1907.

1998 A.D. The ownership of the palimpsest was disputed in federal court in New York in the case of the Greek Orthodox Patriarchate of Jerusalem versus Christie's, Inc. The plaintiff contended that the palimpsest had been stolen from one of its monasteries in the 1920s. Judge Kimba Wood decided in favor of Christie's Auction House on laches grounds.

October 29, 1998 Christie's of New York held a much-publicized auction. The Archimedes Palimpsest was sold for two million dollars to an anonymous collector.

4. Archimedes Manuscripts

- *Sand Reckoner*
  Attempts to remedy for the inadequacies of the Greek numerical notation system by showing how to express the number of grains of sand required to fill the universe in a positional (base 100,000,000) numeral system.
- *Equilibrium of Planes* (two volumes)
  Find the centers of gravity of various plane figures and conics, and establishes the "law of the level".
- *Quadrature of the Parabola*
  Finds the area of any segment of a parabola.
- *Measurement of a Circle*
  Showed that the area constant was one quarter of the circumference constant π, and bound this constant between $3\frac{10}{71}$ and $3\frac{10}{70}$.
- *On the Sphere and the Cylinder* (in two volumes)
  Shows that the surface area of any sphere is $A = 4\pi r^2$, and that the volume of a sphere is $V = \frac{4}{3}\pi r^3$.
- *On Spirals*
  Develops the properties of the tangents to the "spiral of Archimedes", given in polar coordinates as $r = a\theta$.
- *On Conoids and Spheroids*
  Finds the volumes of solids of revolution.
- *On Floating Bodies* (two volumes)
  Find the positions that various solids will assume when floating in a fluid, and establishes "Archimedes' principle" (that the buoyant force on a submerged object is equal to the weight of the displaced fluid).
- *The Method*
  Describes the process of discovery in mathematics.

5. Precursors of Archimedes

5.1. Pythagorean Irrational Numbers. The Pythagoreans proved the existence of irrational numbers in the form of "incommensurable quantities". This tore at the fabric of their world view, based on the supremacy of whole numbers, and it is legend that the demonstrator of irrational numbers was thrown overboard at sea.

5.2. Zeno's Paradoxes. Zeno (ca. 450 B.C.) developed his famous "paradoxes of motion".

5.2.1. *The Dichotomy.* The first asserts the non-existence of motion on the ground that that which is in locomotion must arrive at the half-way stage before it arrives at the goal. (Aristotle Physics, 239b11).

5.2.2. *Achilles and the Tortoise.* The [second] argument was called "Achilles", accordingly, from the fact that Achilles was taken [as a character] in it, and the argument says that it is impossible for him to overtake the tortoise when pursuing it. For in fact it is necessary that what is to overtake [something], before overtaking [it], first reach the limit from which what is fleeing set forth. In [the time in] which what is pursuing arrives at this, what is fleeing will advance a certain interval, even if it is less than that which what is pursuing advanced. And in the time again in which what is pursuing will traverse this [interval] which what is fleeing advanced, in this time again what is fleeing will traverse some amount. And thus in every time in which what is pursuing will traverse the [interval] which what is fleeing, being slower, has already advanced, what is fleeing will also advance some amount. (Simplicius(b) On Aristotle's Physics, 1014.10)

5.2.3. *The Arrow.* The third is that the flying arrow is at rest, which result follows from the assumption that time is composed of moments. He says that if everything when it occupies an equal space is at rest, and if that which is in locomotion is always in a now, the flying arrow is therefore motionless. (Aristotle Physics, 239b.30) Zeno abolishes motion, saying "What is in motion moves neither in the place it is nor in one in which it is not". (Diogenes Laertius Lives of Famous Philosophers, ix.72)

5.2.4. *The Stadium.* The fourth argument is that concerning equal bodies [AA] which move alongside equal bodies in the stadium from opposite directions – the ones from the end of the stadium [CC], the others from the middle [BB] – at equal speeds, in which he thinks it follows that half the time is equal to its double. And it follows that the C has passed all the As and the B half; so that the time is half. And at the same time it follows that the first B has passed all the Cs. (Aristotle Physics, 239b33)

5.3. Eudoxus Method of Exhaustion. Eudoxus (ca. 370 B.C.) is remembered for two major mathematical contributions: the *Theory of Proportion*, which filled the gaps in the Pythagoren theories created by the existence of incommensurable quantities, and the *Method of Exhaustion*, which dealt with Zeno's Paradoxes. This method is based on the proposition: *If from any magnitude there be subtracted a part not less than its half, from the remainder another part not less than its half, and so on, there will at length remain a magnitude less than any preassigned magnitude of the same kind.*

Archimedes credits Eudoxus with applying this method to find that the volume of "any cone is on third part of the cylinder which has the same base with the cone and equal height."

5.4. Euclid's Elements. Euclid of Alexandria (ca. 300 B.C.) wrote *The Elements*, which may be the second most published book in history (after the Bible). The work consists of thirteen books, summarizing much of the basic mathematics of the time, spanning plane and solid geometry, number theory, and irrational numbers.

Among the results in Euclid, we find

Result 1. *The circumferences of two circles are to each other as their diameters.*

Using modern notation, this says that if we are given two circles with diameters D_1 and D_2, and circumferences C_1 and C_2, then

$$\frac{C_1}{C_2} = \frac{D_1}{D_2}.$$

We can rearrange this to say

$$\frac{C_1}{D_1} = \frac{C_2}{D_2}.$$

From this, one may conclude that for any given circle, the ratio between the circumference and the diameter is a constant:

$$\frac{C}{D} = p, \quad \text{so} \quad C = pD.$$

We shall call p the *circumference constant*.

Euclid later shows

Result 2. *The areas of two circles are to each other as the squares of their diameters.*

That is, if A_1 and A_2 represent the area of the circles, then

$$\frac{A_1}{D_1^2} = \frac{A_2}{D_2^2},$$

which says that there is an *area constant* for any circle:

$$\frac{A}{D^2} = k, \quad \text{so} \quad A = kD^2.$$

However, Euclid doesn't mention, and possibly doesn't realize, that p and k are related.

Later still, Euclid shows

Result 3. *The volumes of two spheres are to each other as the cubes of their diameters.*

Thus if V_1 and V_2 are the volumes of spheres of diameter D_1 and D_2, then

$$\frac{V_1}{D_1^3} = \frac{V_2}{D_2^3};$$

again, one sees that, for again given sphere, there is a *volume constant* m such that

$$\frac{V}{D^3} = m, \quad \text{so} \quad V = mD^3.$$

6. Measurement of a Circle

Proposition 4. *The area of any circle is equal to a right-angled triangle in which one of the sides about the right angle is equal to the radius, and the other to the circumference, of the circle.*

Let be C be the circumference, r the radius, and A the area of the circle. Let T be the area of a right triangle with legs of length r and C. Then $T = \frac{1}{2}rC$. Archimedes claims that $A = T$, so $A = \frac{1}{2}rC$.

Lemma 5. *Let h be the apothem and let Q be the perimeter of a regular polygon. Then the area of the polygon is*

$$P = \frac{1}{2}hQ.$$

Proof. Suppose the polygon has n sides, each of length b. Clearly $Q = nb$. Then the area is subdivided into n triangles of base b and height h, so

$$P = n(\frac{1}{2}hb) = \frac{1}{2}hQ.$$

□

Lemma 6. *Consider a circle of area A let $\epsilon > 0$. Then there exists an inscribed polygon with area P_1 and a circumscribed polygon with area P_2 such that*

$$A - \epsilon < P_1 < A < P_2 < A + \epsilon.$$

Proof. Archimedes simply says: "Inscribe a square, then bisect the arcs, then bisect (if necessary) the halvesand so on, until the sides of the inscribed polygonwhose angular points are the points o the division subtend segments whose sum is less than the excess of the area of the circle over the triangle." □

does not explicitly prove this

Proof of Proposition. By double reductio ad absurdum.

Suppose that $A > T$. Then $A - T > 0$, so there exists an inscribed regular polygon with area P such that $A - P < A - T$. Thus $P > T$. If Q is the perimeter and h the apothem of the polygon, we have

$$P = \frac{1}{2}hQ < \frac{1}{2}rC = T,$$

a contradiction.

On the other hand, suppose that $A < T$. Then $T - A > 0$, so there exists a circumscribed polygon with area P such that $P - A < T - A$. Thus $P < T$. However, if Q is the perimeter and h the apothem of the polygon, we have

$$P = \frac{1}{2}hQ > \frac{1}{2}rC = T,$$

a contradiction.

Therefore, as Archimedes writes, "since then the area of the circle is neither greater nor less than [the area of the triangle], it is equal to it." □

Proposition 7. *The ratio of the circumference of any circle to its diameter is less the* $3\frac{1}{7}$ *but greater than* $3\frac{10}{71}$.

Proof. Inscribe a hexagon. Compute the area:

$$\pi = \frac{C}{D} > \frac{Q}{D} = \frac{6r}{2r} = 3.$$

Archimedes next doubles the number of vertices to obtain a regular dodecagon. The computation of its area requires accurate extraction of $\sqrt{3}$, which Archimedes estimates as

$$\frac{265}{153} < \sqrt{3} < \frac{1351}{780},$$

which is impressively close. The Archimedes continues with 24, 48, and finally 96 sides, at each stage extracting more sophisticated square roots.

Next circumscribe a hexagon and continue to 96 sides. □

7. On the Sphere and the Cylinder

The two volume work entitled *On the Sphere and the Cylinder* is Archimedes undisputed masterpiece, probably regarded by Archimedes himself as the apex of his career. These two volumes are constructed in a manner similar to Euclid's Elements, in that it proceeds from basic definitions and assumptions, through simpler known results, onto the new discoveries of Archimedes.

Among the results in this work are the following.

Proposition 8. *The surface of any sphere is equal to four times the greatest circle in it.*

Technique of Proof. Double reductio ad absurdum: assumption that the area is more leads to a contradiction, as does assumption that the area is less. □

Let us translate this into modern notation. Let r be the radius of the sphere and let S be its surface area. Then the radius of the greatest circle in it is πr^2. Thus Archimedes shows that

$$S = 4\pi r^2.$$

Proposition 9. *Any sphere is equal to four times the cone which has its base equal to the greatest circle in the sphere and its height equal to the radius of the sphere.*

Note that again, Archimedes has expressed the volume of the sphere in terms of the volume of a known solid; this is because the Greeks did not have modern algebraic notation. Using modern notation, we let r be the radius of the V be the volume of the sphere. The volume of the cone of radius r and height r, as determined by Eudoxus, is $\frac{1}{3}\pi r^3$. Thus

$$V = \frac{4}{3}\pi r^3.$$

In this way, Archimedes found the relationship between the circumference constant p, the area constant k (in *Measurement of a Circle*), and the volume constant m: We have

$$C = pD, \quad A = kD^2, \quad , \text{ and } V = mD^3,$$

and Archimedes has shown (in modern notation) that

$$C = \pi D \qquad \text{(that is, } p = \pi\text{)}$$

$$A = \pi r^2 = \pi\left(\frac{D}{2}\right)^2 = \frac{\pi}{4}D \qquad \text{(so } k = \frac{\pi}{4}\text{)}$$

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi\left(\frac{D}{2}\right)^3 = \frac{\pi}{6}\pi D^3 \qquad \text{(so } m = \frac{\pi}{6}\text{)}$$

From here, Archimedes now describes an astounding discovery.

Suppose we have a sphere of radius r, surface area S, and volume V. Inscribe this sphere in a right circular cylinder, whose radius would also be r and whose height would be $2r$. Then the surface area A_{cyl} of the cylinder is simply the areas of the base and top circle, plus the area of the rectangle which forms the tube of the cylinder:

$$A_{\text{cyl}} = 2(\pi r^2) + (2\pi r)(2r) = 6\pi r^2.$$

Thus

$$A_{\text{cyl}} : A_{\text{sph}} = (6\pi r^2) : (4\pi r^2) = 3 : 2.$$

Moreover, the volume of the cylinder is the area of the circular base times the height:

$$V_{\text{cyl}} = (\pi r^2)(2r) = 2\pi r^3.$$

Again, we have

$$V_{\text{cyl}} : V_{\text{sph}} = (2\pi r^3) : (\frac{4}{3}\pi r^3) = 3 : 2.$$

This so impressed Archimedes that he requested that his tombstone be engraved with a sphere inscribed in a cylinder, together with the ration 3 : 2. Apparently, Marcellus, the conqueror of Syracuse, was so impressed with Archimedes, that he granted this wish.

8. Equilibrium of Planes

In the treatise *Equilibrium of Planes*, Archimedes establishes the *law of the lever*.

Result 10. *Let W_1 and W_2 be the respective weights of two objects placed on a lever on opposite sides of a fulcrum with respective distances d_1 and d_2. Then the objects balance if and only if*

$$d_1 W_1 = d_2 W_2.$$

This is proven by Archimedes, following three assumptions:

- **(a)** Equal weights at equal distances from the fulcrum balance. Equal weights at unequal distances from the fulcrum do no balance, but the weight at the greater distance will tilt its end of the lever down.
- **(b)** If, when two weights balance, we add something to one of the weights, they no longer balance. The side holding the weight we increased goes down.

(c) If, when two weights balance, we take something away from one, they no longer balance. The side holding the weight we did not change goes down.

9. The Method

An *infintesimal* is a number greater in absolute value than zero, yet smaller than any positive real number. A number $x \neq 0$ is an infintesimal if and only if every sum $|x| + \cdots + |x|$of finitely many terms is less than 1, no matter how large the finite number of terms. In that case, $\frac{1}{x}$ is larger than any positive real number.

Using modern techniques, it is possible to construct a field which contains the real numbers as a subfield, and which contains infinitesimals. However, it is not, and cannot be, complete. This last fact was known to Archimedes, and in fact is called the *Archimedean property* of the real numbers:

Proposition 11 (Archimedean Property)**.** *Let $\mathbb{F}$ be a complete ordered field (for example, $\mathbb{F} = \mathbb{R}$). Let $a, b \in \mathbb{F}$. Then there exists a natural number $n \in \mathbb{N}$ such that $na > b$.*

Proof. Suppose not, then the set $X = \{na \mid n \in \mathbb{N}\}$ is bounded, and by completeness, it has a least upper bound, say c.

ETC. □

Thus, Archimedes worked under the premise that infinitesimals were appropriate for intuitive thought and discovery, but not for proof.

For example, in *Quadrature of a Parabola*, Archimedes uses the method of exhaustion to show that the area of a segment of a parabola is

$$\text{area segment} = \frac{4}{3}\text{area inscribed triangle.}$$

But the method of exhaustion is not how he discovered this formula. In fact, in *The Method*, he writes: *"certain things first became clear to me by a mechanical method, although they had to be proved by geometry afterwards because their investigation by the said method did not furnish an actual proof. But it is of course easier, when we have previously acquired, by the method, some knowledge of the questions, to supply the proof than it is to find it without any previous knowledge."*

As an example of this, consider the first proposition from the palimpsest.

APPENDIX B

Computing

1. What is Computing?

1.1. Definitions. Modern computers receive, store, process, and transmit information. Information, to a computer, consists of a sequence of zeros and ones.

Electrical current transmitted on a wire is typically referred to as either

- analog: continuously varying;
- digital: either on or off.

In digital transmission, with on being 1 and off being 0, we see a sequence of zeros or ones. So, computers understand digital information.

How does the computer interpret the zeros and ones? Why is this called "digital"?

To understand this more fully, first we investigate the meaning of "to compute".

According to my dictionary, we have these definitions:

- Compute: to determine by reckoning; to calculate
- Calculate: to reckon or determine by reasoning
- Reckon: to count

So by definition, a computer counts and uses reasoning (that is, logic). Counting and reasoning combine to produce arithmetic (adding, multiplying, etc.) The word "calculate" comes from *calculus*, a pebble used in counting. This in turn comes from *calx*, or limestone.

1.2. Counting. There is archeological evidence that man began counting as far back as 50,000 years ago.

The jaw bone of a wolf has been discovered which is 20,000 years old, and has 25 notches slashed in groups of five. So, this bone is evidence that man used "technology" to aid in counting at least that long ago.

However, this evidence indicates that actually, an earlier counting technology had been developed: that man counted on his five fingers per hand earlier than this.

Many undeveloped tribes have been discovered that use the word "hand" to mean five. A "man" may mean ten or twenty, depending on the tribe. One tribe used the word "mattress" to mean forty.

Now when we count to ten on two hands, the number represented is simply the number of fingers held up. Which fingers, on which hand, is irrelevant to the number indicated.

However, consider counting to thirty on two hands by letting the fingers on the left hand each have a value of five. When the right hand reaches five fingers

up, lower them and raise an additional finger on the left hand. This is counting in *base five*.

The ancient Mayans had a method of counting uses rocks and sticks. Each rock was worth one and each stick was worth five. Then, one could trade five rocks for a stick. This later became how they wrote numbers. After four sticks, however, the Mayans got tired. What could they do? They used position to indicate powers of twenty. Under this scheme, each position represented a power of twenty, and a shell was used as a place holder (that is, a shell is zero).

Actually, we can count from 0 to 31 on one hand. Why 31? Because it is $2^5 - 1$. Here's how. Evaluate each finger on your right hand as follows:

- right thumb = 1
- right index finger = 2
- right middle finger = 4
- right ring finger = 8
- right pinky = 16

Then count according to this chart:

0	no fingers	16	pinky
1	thumb	17	pinky + thumb
2	index	18	pinky + index
3	index + thumb	19	pinky + index + thumb
4	middle	20	pinky + middle
5	middle + thumb	21	pinky + middle + thumb
6	middle + index	22	pinky + middle + index
7	middle + index + thumb	23	pinky + middle + index + thumb
8	ring	24	pinky + ring
9	ring + thumb	25	pinky + ring + thumb
10	ring + index	26	pinky + ring + index
11	ring + index + thumb	27	pinky + ring + index + thumb
12	ring + middle	28	pinky + ring + middle
13	ring + middle + thumb	29	pinky + ring + middle + thumb
14	ring + middle + index	30	pinky + ring + middle + index
15	ring + middle + index + thumb	31	all fingers

This is counting using a positional base two scheme. Positional, because the position of the finger which is up or down matters. Base two, because each finger has two possible values (up or down).

Suppose we evaluate each finger on our left hand as follows:

- left thumb = 32
- left index finger = 64
- left middle finger = 128
- left ring finger = 256
- left pinky = 512

Then we could count to $2^{10} - 1 = 1023$. If we additionally used our toes, we could count to $2^{20} - 1 = 1048575$.

1.3. Digital Information. Computers use only zeros and ones to count, reason, and store information.

We have seen how two count using positioned zeros and ones, and we will explore this more later in our study of bases.

Information is received, stored, and transmitted as a sequence of zeros and ones. The program must interpret how to view a given sequence of zeros and ones; this is done through the concept of *data types*. Having gotten used to binary numbers through studying bases, we will then explore how computer store different types of information in binary codes.

Zeros and ones are also used as the logical values of *false* and *true*. Logical operators such as AND and OR combine inputs of zeros and ones to create an output. We will study the *truth tables* which define the logical operators, and then briefly look at the circuitry which implements these operators through what are known as *gates*. This will give some idea of how a computer reasons.

2. A Brief History of Early Computing

2.1. Definitions.

- Compute: to calculate.
- Calculate: to reckon; to reason.
- Reckon: to count.
- Reason: to think logically.
- The word calculate is derived from calculus, meaning a stone used in counting, which is in turn derived from calx, meaning limestone.

2.2. Primitive Counting Terminology.

- Digit: one (one finger).
- Hand: five (five fingers).
- Man: ten (ten fingers).
- Man: twenty (ten fingers, ten toes).
- Mattress: forty.

2.3. Prehistoric Computing.

- 50,000 years old: archeological evidence of counting by man.
- 20,000 years old: bone with notches in groups of five.

2.4. Sticks and Stones.

- Ancients Mayans used sticks and stones to count. A stone was worth one and a stick was worth five.
- A number consisted of a group of stick and stones; adding the values of the sticks and stones produced the number. Adding two numbers amounted to combining the two groups.
- To represent large numbers, Mayans developed a positional base twenty system wherein a shell represented zero.

2.5. Mayan Numerals.

- The Mayans later developed writing, and their symbols for numerals reflected these origins.

2.6. Egyptian Numerals.

- The Egyptians used hieroglyphic numerals in an additive, nonpositional system. These pictures were carved in stone.

2.7. Babylonian Numerals.

- The Babylonians produced cuneiform writing, which consisted of wedges pressed into clay tablets. They developed a base sixty positional numeral system.

2.8. Sand Trays (2400 B.C.).

- The ancient Babylonians used sand trays to do mathematical scratch work.
- Combining the sand tray with stones led to a computational tool, using stones to represent one, ten, sixty, etc., depending on their position.

2.9. Counting Boards (300 B.C.).

- Boards with fixed positions were designed to hold the stones used in computation.
- The Salamis tablet is a counting board used by the Babylonians circa 300 B.C. It is a slab of white marble measuring 149cm in length, 75cm in width and 4.5cm thick.

2.10. The Abacus (100 A.D.).

- Eventually, the stones were place on rods to fix their position. This led to the abacus, from the Greek word abax, meaning sand tray.

2.11. Origins of Algebra (800 A.D.).

- Ancient Greeks, masters of geometry, had no algebra and a difficult numeral system.
- Ancient Hindus invented our current numeral system, using positional base ten and zero.
- Hindus and Arabs explored algebraic notation.
- "Algebra" comes from the Arabic al-jabr "reunion", "resetting of broken parts", used in the title of al-Khwarizmi's influential work ilm al-jabr wa'l-muqbala, "the science of restoration and equating like with like".

2.12. Middle Ages (300-1200 A.D.).

- Technology, knowledge, and nearly all intellectual endeavors came to a virtual standstill in Europe during the period of dominance by the Holy Roman Empire.
- Communication with the east sparked the European Renaissance, and with it, ideas for computational technology.

2.13. Da Vinci's Mechanical Calculator (1500 A.D.).

- Leonardo Da Vinci, Italian painter, musician, sculptor, architect, and engineer, created drawings of a mechanical calculator, working models of which have since been constructed.

2.14. Napier's Bones (1600 A.D.).

- John Napier, Scottish mathematician, inventor of logarithms, invented a tool called Napier's Bones, which were multiplication tables inscribed on strips of wood or bone.

2.15. Oughtred's Slide Rule (1620 A.D.).

- William Oughtred, English mathematician and clergyman, early explorer of Calculus, invented the slide rule using Napier's logarithms.

2.16. Pascal's Arithmetic Machine (1640 A.D.).

- Blaise Pascal, French mathematician, physicist, and theologian, is credited with the invention of the first operational calculating machine. He developed an operating model of the Arithmetic Machine to help his father add sums of money.

2.17. Leibnitz' Step Reckoner (1670 A.D.).

- Gottfried von Leibnitz: French mathematician, philosopher, and lawyer, cocreator of Calculus, developed the Step Reckoner, a device which, as well as performing additions and subtractions, could multiply, divide, and evaluate square roots by series of stepped additions.

2.18. Jacquard's Punched Cards (1800 A.D.).

- Joseph-Marie Jacquard, French silk weaver, invented a way of automatically controlling the warp and weft threads on a silk loom by recording patterns of holes in a string of cards.

2.19. Babbage's Engines (1830 A.D.).

- Charles Babbage, English mathematician and inventor, designed the Difference Engine to automatically compute mathematical tables.
- Later, he designed the Analytical Engine, intended to use punched cards, sequencing, branching, and looping to control an automatic calculator.

2.20. Wheatstone and Morse's Telegraphs (1840 A.D.).

- Sir Charles Wheatstone invented the first British telegraph.
- Samuel Morse invented the first American telegraph, using dots and dashes. This became the standard known as Morse code.

2.21. Wheatstone's Tape (1860 A.D.).

- Wheatstone introduced paper tapes as a medium for the preparation, storage, and transmission of data in the form of Morse Code.
- The paper tape used two rows of holes to represent Morse's dots and dashes. Outgoing messages could be prepared off-line on paper tape and transmitted later.

2.22. Sholes' Keyboard (1874 A.D.).

- Christopher Latham Sholes invented the QWERTY keyboard.

2.23. Hollerith's Tabulating Machine (1890 A.D.).

- American inventor Herman Hollerith used punched cards to represent the data gathered for the 1890 American census, and to read and collate this data using an automatic machine. His company became IBM in 1924.

2.24. De Forest's Vacuum Tubes (1906 A.D.).

- In 1879, Thomas Edison demonstrates the incandescent light bulb.
- In 1883, John Ambrose Fleming uses this to convert electromagnetic radiation into electricity, the precursor of the radio. He produced 2-element vacuum tubes (diodes).
- In 1906, Lee de Forest produced 3-element vacuum tubes (triodes), which could be used as both an amplifier and a switch.

2.25. Turing's Machine (1937 A.D.).

- Alan Turing, English logician and mathematician, invented the abstract Turing Machine, a theoretical construct which helped prove the non-computability of certain arithmetic results.
- The abstraction involves the movement of a sequence of cells called a "tape".

2.26. Flower's COLOSSUS (1941 A.D.).

- Sir Tommy Flowers, British engineer, together with Turing, designed and constructed COLOSSUS from 1941 to 1943 during WW II to break German encryption. COLOSSUS has been credited as the 1st programmable computer.

2.27. Princeton's ENIAC (1943 A.D.).

- Princeton University constructed ENIAC, the 1st truly general purpose programmable computer, between 1943 and 1946.
- Miles of wiring
- 70,000 resistors, 10,000 capacitors
- 18,000 vacuum tubes
- No monitor
- 3,000 blinking lights
- Cost 486,000 dollars
- 100,000 additions per second
- Weighed 30 tons
- Filled a 30x50 foot room
- Could be replaced today by one fingernail-size silicon chip

2.28. The First "Bug" (1945 A.D.).

- On September 9th, 1945, a moth flew into the wiring of the Harvard Mark II Relay Calculator, which was in service at the Naval Weapons Center in Dahlgren, Virginia, causing it to malfunction.
- The operator "debugged" the machine by removing the insect.

2.29. Eckert-Mauchly's UNIVAC (1951 A.D.).

- John William Mauchly and J. Presper Eckert Jr. produce the UNIVAC (Universal Automatic Computer), the 1st commercially available computer.
- Handled letters as well as numbers.
- Separated input/output from computation.
- The Eckert-Mauchly Computer Company was eventually purchased by Sperry-Rand.

2.30. Transistors.

- Transistors, which replaced vacuum tubes, were developed between 1947 and 1959.
- The advent of transistors revolutionized electronics and computing; this is a story unto itself, to be explored in the next installment.

2.31. References.

- http://www.maxmon.com/history.htm
- http://www.pbs.org/transistor/index.html

APPENDIX C

Important Mathematicians

1. Obscured by Time

Prehistoric

Egyptian

Babylonian

2. Ancient Greek Geometry

Thales of Miletus (Greek 624 BC - 547 BC)

Pythagoras of Samos (Greek 569 BC - 475 BC)

Hippocrates of Chios (Greek 470 BC - 410 BC)

Plato of Athens (Greek 427 BC - 347 BC)

Eudoxus of Cnidus (Greek 408 BC - 355 BC)

Euclid of Alexandria (Greek 325 BC - 265 BC)

Archimedes of Syracuse (Greek 287 BC - 212 BC)

Apollonius of Perga (Greek 262 BC - 190 BC)

Diophantus of Alexandria (Greek 200 AD - 284 AD)

3. Ancient Greek Astronomy

Aristotle of Athens (Greek 384 BC - 322 BC)

Aristarchus of Samos (Greek 310 BC - 230 BC)

Eratosthenes of Cyrene (Greek 276 BC - 194 BC)

Ptolemy of Alexandria (Roman 85 - 165)

4. Transition

Sun Zi (Chinese 400-460)

Abu Ja'far Muhammad ibn Musa Al-Khwarizmi (Arabic 790 - 840)

Omar Khayyam (Persian 1048 - 1122)

Leonardo Pisano Fibonacci (Italian 1170 - 1250)

5. Cubic Polynomials

Luca Pacioli (Italian 1445 - 1517)

Scipione del Ferro (Italian 1465 - 1526)

Nicolo Fontana Tartaglia (Italian 1500 - 1557)

Girolamo Cardano (Italian 1501 - 1576)

Lodovico Ferrari (Italian 1522 - 1565)

Rafael Bombelli (Italian 1526 - 1572)

Franois Vite (French 1540 - 1603)

6. Logarithms

Johann Werner (German 1468 - 1522)

John Napier (Scottish 1550 - 1617)

Henry Briggs (English 1561 - 1630)

7. Early Astronomy

Leonardo da Vinci (Italian 1452 - 1519)

Nicolaus Copernicus (Polish 1473 - 1543)

Tycho Brahe (Danish 1546-1601)

Galileo Galilei (Italian 1564 - 1642)

Johannes Kepler (German 1571 - 1630)

8. Analytic Geometry

René Descartes (French 1596 - 1650)

Pierre de Fermat (French 1601 - 1665)

Blaise Pascal (French 1623 - 1662)

9. Early Calculus

Bonaventura Cavalieri (Italian 1598-1647)

John Wallis (English 1619-1703)

Isaac Barrow (English 1630 - 1677)

Isaac Newton (English 1642-1727)

Gottfried Leibniz (German 1646-1716)

APPENDIX D

Problems

In construction problems, describe each step, and draw all steps with a straight-edge and compass, labeling each point significant for the construction. Explain why your construction works.

1. Easier

Problem 1. State the base and type (simple, multiplicative, ciphered, or positional) used by the given numeral system.

(a) Egyptian hieroglyphic
(b) Babylonian
(c) Greek
(d) Chinese
(e) Mayan

Problem 2. Write 1070 in Mayan.

Problem 3. Find the base six radix expansion of $\frac{271}{54}$.

Problem 4. (Babylonian Fractions)
Let $x = \frac{5}{72}$ and $b = 60$. Find the base b radix expansion of x. (Hint: $x = \frac{5a}{72a}$, where $72a$ is a power of 60.)

Problem 5. Using a straight edge and compass, construct the circle passing through three given points. Label the original points, all necessary constructed points, and describe exactly how the constructed points were created.

Problem 6. (Greek Geometry)
Find the area of a regular octagon inscribed in a unit circle.

Problem 7. The key definition of Eudoxus' theory of proportion is laid out in Euclid's Elements.

Elements Book V Definition 5
Magnitudes are said to be in the same *ratio*, the first to the second and the third to the fourth, when, if any *equimultiples* whatever are taken of the first and third, and any equimultiples whatever of the second and fourth, the former equimultiples *alike exceed*, are *alike equal to*, or alike *fall short of*, the latter equimultiples respectively taken in corresponding order.

Let $a, b, c, d \in \mathbb{R}$ be positive real numbers. Consider the proposition:

$$\frac{a}{b} = \frac{c}{d} \quad \Leftrightarrow \quad \forall m, n \in \mathbb{N}, \begin{cases} ma > nb \Leftrightarrow mc > nd; \\ ma = nb \Leftrightarrow mc = nd; \\ ma < nb \Leftrightarrow mc < nd. \end{cases}$$

(a) Identify the italicized words in the definition with the exact mathematical symbols in the proposition.
(b) Prove the proposition from a modern perspective.

Problem 8. (Euclidean Algorithm)
Let $m = 80$ and $n = 167$. Find $x, y, d \in \mathbb{Z}$ such that $d = \gcd(m, n)$ and

$$mx + ny = d.$$

Problem 9. (Diophantus' Theorem)
Let $a, b, c \in \mathbb{Z}$ with $a^2 + b^2 = c^2$. Show that if a is odd, then $b + c$ is a perfect square.

Problem 10. Find the area of a regular octagon inscribed in the unit circle.

Problem 11. Consider the cubic curve with equation

$$y^2 = x^3 - 3x + 1.$$

Find a rational point on the curve other than $(0, \pm 1)$.

Problem 12. Let $n \in \mathbb{Z}$ with $n \geq 2$. Let $a, b, c, d \in \mathbb{Z}$ with $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$.
Show that $ab \equiv cd \pmod{n}$.

Problem 13. Find $c \in \mathbb{Z}$ with $0 \leq c < 221$ such that $c \equiv 7 \pmod{13}$ and $c \equiv 11 \pmod{17}$.

Problem 14. Let $m = 41$, $n = 61$, $a = 21$, and $b = 31$.

(a) Find x and y so that $mx + ny = 1$.
(b) Find $c \in \mathbb{Z}$ with $0 \leq c < 3501$ such that $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$.

Problem 15. Define a sequence for real number (G_n) by $G_1 = 1$, $G_2 = 1$, and $G_{n+2} = 3G_n + G_{n+1}$.
Let (a_n) be the sequence defined by $a_n = \frac{G_{n+1}}{G_n}$.

(a) Compute the first 5 terms of (G_n).
(b) Compute the first 5 terms of (a_n).
(c) Write a_{n+1} in terms of a_n.
(c) Compute $\lim a_n$.

Problem 16. Consider the cubic equation

$$x^3 + 3x^2 + 6x + 7 = 0.$$

(a) Substitute $x = y - 1$ to obtain an equation without a y^2 term.
(b) Use the method of Tartaglia to compute y which satisfies this equation.
(c) Find x which satisfies the original equation.

Problem 17. (Titles)
Indicate the author of each manuscript. Choose from these authors:
Archimedes, Appolonius, Cavalieri, Descartes, Diophantus, Euclid, Fibonacci, Gauss, Napier, Newton.

- **(a)** *Principia Mathematica*
- **(b)** *Liber Abaci*
- **(c)** *Arithmetica*
- **(d)** *The Elements*
- **(e)** *La geometrie*
- **(f)** *Disquisitiones arithmeticae*
- **(g)** *Conic Sections*
- **(h)** *Geometrica indivisibilibus*
- **(i)** *On the Sphere and the Cylinder*

Problem 18. (Archimedes)
To compute π, Archimedes found the areas of many regular polygons. Find the area of a regular dodecagon inscribed in the unit circle.

Problem 19. (Diophantus)
To find Pythagorean triples, Diophantus consider the intersection of the unit circle with a line through $(-1, 0)$ with rational slope. Find the Pythagorean triple (a, b, c), with $a, b, c \in \mathbb{Z}$, $\gcd(a, b, c) = 1$, and $a^2 + b^2 = c^2$, that this technique produces when the slope of the line is $\frac{3}{5}$.

Problem 20. (Tartaglia)
To solve equations of the form $x^3 + mx = n$, Tartaglia set $x = t - u$ and used the substitutions $m = 3tu$ and $n = t^3 - u^3$. Apply this technique to find the solutions to $x^3 + 9x = 20$.

Problem 21. (Descartes)
To compute tangents, Descartes used the discriminant of a quadratic equation to find the circle centered on the x-axis and tangent to a given curve at a given point. Use this technique to find the center $(a, 0)$ of such a circle, when the equation is $2x^2 - y^2 = 1$, the point is $(1, 1)$, and $a > \frac{1}{2}$.

Problem 22. (Leibnitz)
To compute the sum of the reciprocals of the triangular numbers, Leibnitz used a telescoping sum. Reproduce this argument.

Problem 23. (Euler)
To compute the sum of the reciprocals of the square numbers, Euler considered the zeros of the function $\sin x/x$ to produce the equation

$$\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k+1)!} = \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{\pi^2 n^2}\right).$$

He then equated the coefficients of the x^2 term on both sides and obtained $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

- **(a)** Compute the sum of the reciprocals of the even square numbers by substituting $x \mapsto x/2$.
- **(b)** Compute the sum of the reciprocals of the odd square numbers by subtraction.

Problem 24. (Pythagorean Triples)
The Babylonians generated tables of Pythagorean triples (a, b, c) such that a is sexagesimally regular. Euclid's *Elements* supplied a technique for computing Pythagorean triples using the equations

$$a = 2uv, \quad b = u^2 - v^2, \quad c = u^2 + v^2.$$

Diophantus proved that this produces *all* Pythagorean triples.

Thus the following function generates Pythagorean triples:

$$\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N} \times \mathbb{N} \quad \text{by} \quad \phi(u, v) = (2uv, u^2 - v^2, u^2 + v^2).$$

Set

$$S = \{n \in \mathbb{N} \mid \ 1 \leq n \leq 10 \text{ and } n \text{ is decimally regular}\};$$
$$U = \{(u, v) \in S \times S \mid \ v < u \text{ and } \gcd(u, v) = 1\}.$$

(a) Find S.
(b) Find U.
(c) Find $\phi(U)$.

Problem 25. (Regular Solids)
The *regular solids* were studied by the Pythagoreans, the Platonists, and Euclid.

(a) List the regular solids. State the type of regular polygon from which each solid is constructed. Find the number of faces F, the number of edges E, and the number of vertices V. Compute $F - E + V$.
(b) Luca Pacioli (1509) used three intersecting golden rectangles to construct a regular solid whose faces are equilateral triangles with sides of length one. Use this construction to find the radius of a sphere in which such a solid can be transcribed.

Problem 26. (Diophantine Geometry)
A *rational curve* is the set of solutions to a polynomial equation in two variables whose coefficients are rational numbers. A *rational point* on a curve is a solution whose coordinates are rational numbers.

Diophantus (Alexandria, 2nd century A.D.) realized that, given two rational points on a cubic curve, the slope between them would be rational, and so the third point of intersection between the line and the curve would produce another rational point.

Consider the curve given by the equation

$$y^2 = x^3 - 4x + 9.$$

By trying small values for x, find four rational points on this curve. Select two points such that the slope of the line between them is 3. Compute this line. Intersect this line with the curve to find two additional rational points.

Problem 27. (Congruence)
Euclid's *Elements* contains a description of the Euclidean algorithm for find x, y such that

$$mx + ny = \gcd(m, n).$$

The proof of the *Chinese Remainder Theorem* uses this fact to produce solutions to systems of congruences of the form

$$a \equiv c \pmod{m};$$
$$b \equiv c \pmod{n}.$$

Let $m = 17$, $n = 37$, $a = 7$, and $b = 11$.

- **(a)** Find x and y such that $mx + ny = 1$.
- **(b)** Find c with $0 \leq c < mn$ such that $a \equiv c \pmod{m}$ and $b \equiv c \pmod{n}$.

Problem 28. Constructibility [Extra Credit]
Let A be a set of points in a plane $\mathcal{P}$. Let $\mathcal{L}(A)$ be the set of all lines in $\mathcal{P}$ which pass through at least two points in A, and let $\mathcal{C}(A)$ be the set of all circles in $\mathcal{P}$ pass through a point in A and whose center is a different point in A. Let $\mathcal{O}(A) = \mathcal{L}(A) \cup \mathcal{C}(A)$. Define

$$S(A) = \{z \in \mathcal{P} \mid z \in O_1 \cap O_2 \text{ for some } O_1, O_2 \in \mathcal{O}(A)\}.$$

- **(a)** If A contains one point, how large is $S(A)$?
- **(b)** If A contains two points, how large is $S(A)$?
- **(c)** If A contains three collinear equally spaced points, how large is $S(A)$?
- **(d)** If A contains three collinear unequally spaced points, how large is $S(A)$?
- **(e)** If A contains the vertices of an equilateral triangle, how large is $S(A)$?
- **(f)** If A contains the vertices of an acute isosceles triangle, how large is $S(A)$?
- **(g)** If A contains the vertices of an obtuse isosceles triangle, how large is $S(A)$?

Include a drawing to justify each case.

Problem 29. Given two points A, B in a plane, describe all steps necessary to construct a point C such that $AC \perp AB$ and $\triangle ABC$ is an isosceles triangle.

Problem 30. Let $d = \gcd(728, 231)$. Use that Euclidean Algorithm to find $d, x, y \in \mathbb{Z}$ such that

$$mx + ny = d.$$

2. Harder

Problem 31. Let m and n be integers with $m, n \geq 3$. Let $d = \gcd(m, n)$ and let $k = \frac{mn}{d}$. Given a regular m-gon and a regular n-gon, construct a regular k-gon.

Problem 32 (Regarding Archimedes)**.** Let P be a regular n-gon and let O be its center. Let A and B be consecutive vertices on P and and assume that $|OA| = 1$. Let M be the midpoint between A and B. Find $|OM|$ as a function of n.

Problem 33. Solve the following equations for the positive integers n and b.

- **(a)** $n = (13425)_b = (4115)_{2b}$
- **(b)** $n = (1234)_b = (532)_{2b-1}$

(See Eves Problem Study 1.8.)

Problem 34. A *Pythagorean triple* is and ordered triple (a, b, c) of positive integers such that $a^2 + b^2 = c^2$.

(a) Show that there exists a Pythagorean triple (a, b, c) for every integer $a \geq 3$.
(b) Show that there exist only finitely many Pythagorean triples (a, b, c) for each integer $a \geq 3$.

(See Eves Problem Study 3.6 and discussion on pp. 81-82)

Problem 35. Given line segment $\overline{AB}$ of length 11 and $\overline{CD}$ of length 3, construct a point C on $\overline{AB}$ such that $|CB| = x$, where x is a solution to the quadratic equation

$$x^2 - 11x + 9 = 0.$$

State the exact value of x. (See Eves Problem Study 3.10a and discussion on pp. 88-89)

Problem 36. Given a two points A and B, construct a point Z such that $\angle BAZ = \angle ABZ = 75°$.

Problem 37. Compute the area of a regular pentagon inscribed in a unit circle.

Problem 38. Draw neatly with straightedge and compass, describing each step.
(a) Given two points, construct an angle of 45°.
(b) Trisect the 45° angle.
(c) Does this show that all angles can by trisected?

Problem 39. Using straightedge and compass, construct an angle of 54°. Describe each step, discussing why your construction is effective.

Problem 40. Compute the volume of a regular icosahedron inscribed in a sphere of radius 1.

Definition 1. Let $m, n \in \mathbb{Z}$. The *least common multiple* of m and n is a positive integer $l \in \mathbb{Z}$ such that
(a) $m \mid l$ and $n \mid l$;
(b) $m \mid k$ and $n \mid k$ implies $l \mid k$.

Definition 2. Let $n \in \mathbb{Z}$ with $n \geq 2$. Set $\mathbb{Z}_n = \{r \in \mathbb{Z} \mid 0 \leq r < n\}$. Define a function

$$\rho_n : \mathbb{Z} \to \mathbb{Z}_n \quad \text{by} \quad \rho_n(a) = \text{ the remainder when } n \text{ is divided by } a.$$

We call ρ the *residue map.*

Definition 3. Let $m, n \in \mathbb{Z}$ with $m \geq 2$, $n \geq 2$. Define a function

$$\sigma_{m,n} : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n \quad \text{by} \quad \sigma_{m,n}(a) = (\rho_m(a), \rho_n(a)).$$

We call σ the *joint residue map.*

Problem 41. Let $m, n \in \mathbb{Z}$ with $m \geq 2$ and $n \geq 2$. Let $d = \gcd(m, n)$ and $l = \operatorname{lcm}(m, n)$.
(a) Show that if $d = 1$, then $\sigma_{m,n}$ is bijective.
(b) Show that if $a \equiv b \pmod{l}$, then $\sigma_{m,n}(a) = \sigma_{m,n}(b)$.

Problem 42. (Fibonacci)
Recall that the Fibonacci sequence (F_n) is defined by $F_1 = 1$, $F_2 = 1$, and $F_{n+2} = F_n + F_{n+1}$, and that $\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \phi$, where $\phi = \frac{1+\sqrt{5}}{2}$.

Let $b \in \mathbb{R}$ with $b \geq 1$ and define a sequence (G_n) by $G_1 = 1$, $G_2 = 1$, and $G_{n+2} = G_n + bG_{n+1}$.

Let $c \in \mathbb{R}$ with $c \geq \phi$. Find b such that $\lim_{n\to\infty} \frac{G_{n+1}}{G_n} = c$.

Problem 43. (Tartaglia)
Recall that Tartaglia viewed the cube x^3 as $(t-u)^3$ to find solutions to cubic equations.

Let $f(x) = x^3 + 3x^2 + 6x - 8$. Find the real zero of f using Tartaglia's cube plus cosa method.

Problem 44. (Descartes)
Recall that Descartes used the concept of expanding circles and the ability to compute the number of real solutions to quadratic equations to find tangents.

Find the distance between the curve $x = y^2$ and the point $(3, 0)$ using Descartes' discriminant method.

Problem 45. (Napier)
Recall that Napier desired to find a function to convert multiplication into addition. We may use techniques of Calculus unavailable to him to see that he had very little choice. The modern definition is

$$\log x = \int_1^x \frac{dt}{t} \quad \text{and} \quad \log_b(x) = \frac{\log x}{\log b}.$$

Let $f : (0, \infty) \to \mathbb{R}$ be a differentiable function which is not constantly zero and satisfies

$$f(ab) = f(a) + f(b) \quad \text{for all } a, b \in (0, \infty).$$

Show that there exists $b \in \mathbb{R}$ such that $f(x) = \log_b(x)$.

Problem 46. Let $x = \frac{271}{200}$ (expressed in decimal). Find the base sixty radix expansion of x. (Hint: first multiply the numerator and denominator by some number n, then convert the numerator to base six. If you choose n wisely, you will now be almost done.)

Problem 47. Given two points A and B, construct a point C so that $\triangle ABC$ has angles $30°$, $60°$, and $90°$.

Problem 48. Let $m = 52$, $n = 77$, $a = 5$, and $b = 7$.

- **(a)** Find $x, y \in \mathbb{Z}$ such that $mx + ny = 1$.
- **(b)** Find $c \in \mathbb{Z}$ with $0 \leq c < mn$ such that $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$.

Problem 49. Consider $\mathbb{Z}_{31} = \{0, 1, \ldots, 30\}$ (the bar notation is understood).

- **(a)** Find $a \in \mathbb{Z}_{31}$ such that $2a = 1$.
- **(b)** Find $c, d \in \mathbb{Z}_{31}$ such that $c^2 = d^2 = 5$.
- **(c)** Let $f(x) = x^2 - x - 1$. Find $m, n \in \mathbb{Z}_{31}$ which are distinct solutions to $f(x) = 0$.

Problem 50. Regarding power series:

- **(a)** Write the Taylor series for e^x.
- **(b)** Use the Taylor series of e^x to find the Taylor series of e^{x^2}.
- **(c)** Use the first four terms of the Taylor series of e^{x^2} to estimate $\int_0^2 e^{x^2}\, dx$.

Lemma 4 (Cyclotomic Lemma)**.** *If p be a positive prime integer, then the polynomial*

$$f(X) = 1 + X + X^2 + \cdots + X^{p-1}$$

is irreducible over $\mathbb{Q}$.

Problem 51. Regarding complex numbers and constructibility:

- **(a)** Describe the relationship between a regular n-gon and the zeros of the polynomial $X^n - 1$.
- **(b)** Use the Cyclotomic Lemma to show that a regular heptagon ($n = 7$) is not constructible.

APPENDIX E

Solutions to Problems

Problem 1. (6.7 Apollonius on Tangencies) In his lost treatise on *Tangencies*, Apollonius considered the problem of drawing a circle tangent to three circles, including degenerate forms of a circle including a point or a line.

- **(a)** Find the number of cases, depending on whether we have points, lines, or circles, and the maximum number of solutions in each case.
- **(b)** Given points A and B and line L, find all circles through the points and tangent to the line.
- **(c)** Reduce the case of two lines and a point to the case of part **(b)**.

We will use the following geometric lemmas.

Lemma 1. *Let T be a point on a circle C with center D, and let L be a line through T. Then L is tangent to C if and only if the line through D and T is perpendicular to L.*

Lemma 2. *Let A, B, and C be distinct points. Then there exists a unique circle through A, B, and C whose center is the intersection of the perpendicular bisectors of the line segments $\overline{AB}$, $\overline{BC}$, and $\overline{AC}$.*

Proof. The set of points equally distant between two given points is the line perpendicular to the line through the given points which passes through their midpoint. Thus the center of the circle lies on this line, for each pair of points. □

Lemma 3. (Central Angle Theorem)
Let A, B, and C be points on a circle with center D. Then $\angle ADB = 2\angle ACB$.

Proof. The triangles $\triangle ADB$, $\triangle BDC$, and $\triangle CDA$ are isosceles. Let $y = ACB$ and $2x = \angle ADB$. Now

$$\begin{aligned} 180^\circ &= \angle CAB + \angle CBA + \angle ACB \\ &= (\angle CAD + \angle DAB) + (\angle CBD + \angle DBA) + (\angle ACD + \angle DCB) \\ &= (\angle DAB + \angle DBA) + (\angle CBD + \angle CAD) + (\angle ACD + \angle DCB) \\ &= (\angle DAB + \angle DBA) + 2(\angle ACD + \angle DCB) \\ &= (180^\circ - 2x) + 2y. \end{aligned}$$

Thus $2y - 2x$, so $y = x$. □

Lemma 4. *Let A and B be points and let M be the line through A and B. Let L be a line which is not parallel to M and let S be the point of intersection of L and M. Let T be a point on L. Then L is tangent to the circle through A, B, and T if and only if $\angle STA = \angle SBT$.*

Proof. Let $x = \angle STA$. Let N be the line through T and D. Then

$$\begin{aligned} C \text{ is tangent to } L &\Leftrightarrow L \perp N \\ &\Leftrightarrow \angle STD = 90^\circ \\ &\Leftrightarrow \angle ATD = 90^\circ - x \\ &\Leftrightarrow \angle ADT = 2x \\ &\Leftrightarrow \angle ABT = x. \end{aligned}$$

□

Lemma 5. *Let A and B be points and let M be the line through A and B. Let L be a line which is not parallel to M and let S be the point of intersection of L and M. Let T be a point on L. Then L is tangent to the circle through A, B, and T if and only if*

$$(SA)(SB) = (ST)^2.$$

Proof. By the previous lemma, L is tangent to the circle if and only if $\triangle AST \sim \triangle TSB$, which is true if and only if

$$\frac{AS}{ST} = \frac{ST}{SB}.$$

The result follows. □

Solution.

(a) Letting p mean point, l mean line, and c mean circle, there are ten cases:

$$ppp, ppl, pll, lll, ppc, pcc, ccc, llc, lcc, plc.$$

Type *ppp* has a unique solution; each of the other types has two solution in general.

Two find the unique circle through three points, we take the center to be the intersection of the perpendicular bisectors of the line segments between the points. Thus, the other problems may be reduced to finding the points of tangency on the given lines or circles.

(b) If the line through A and B is parallel to C, then there is a unique solution. The point of tangency on C is obtained by intersecting C with the perpendicular bisector of $\overline{AB}$.

Otherwise, let S be the intersection of the line through A and B and the line C. Let T be a point on C such that $(ST)^2 = (SA)(SB)$. Then T is a point of tangency. There are two solutions (one on either side of S). The proof that this is so follows.

(c) Let A be a point and let L and M be lines.

Let N be the line which bisects the angle between L and M. If L and M happen to be parallel, let N be the midline. Reflect A through this line to obtain a point B. This reduces this case to the previous case, except if A is on N.

If A is on N, construct the line through A perpendicular to N, and let E be the point of intersection. Bisect the angle at E and let D be the point of intersection of the bisecting line with N. The D is the center of the tangent circle

(their are two solutions). To see this, construct the line through D perpendicular to L and intersect it with L at point T. Then $\triangle AED \cong \triangle TED$ by AAS, so $AD = TD$; this is the radius of the circle. $\square$

Problem 2. (6.15 Diophantus)

(a) About all we know of Diophantus' personal life is contained in the following summary of an epitaph given in the *Greek Anthology*: "Diophantus passed $\frac{1}{6}$ of his life in childhood, $\frac{1}{12}$ in youth, and $\frac{1}{7}$ more as a bachelor. Five years after his marriage was born a son who died 4 years before his father, at $\frac{1}{2}$ his father's [final] age." How old was Diophantus when he died?

(b) Solve the following problem, which appears in Diophantus' *Arithmetica* (Problem 17, Book I): Find 4 numbers, the sum of every arrangement 3 at a time being given; say 22, 24, 27, 20.

(c) Solve the following problem, also found in the *Arithmetic* (Problem 16, Book VI): In the right triangle ABC, right angled at C, AD bisects angle A. Find the set of smallest integers for AB, AD, AC, BD, DC such that $DC : CA : AD = 3 : 4 : 5$.

Solution.

(a) Let x be the age of Diophantus at his death, and let y be the number of years he lived after his marriage. Thus $y = x - (\frac{1}{6} + \frac{1}{12} + \frac{1}{7})x$. That is, $y = \frac{17}{28}x$. Also, the final age of his son is $\frac{1}{2}x = (y - 5 - 4) = \frac{17}{28}x - 9$. Solving for x gives $x = 84$.

(b) We could create a system of four linear equations in four variables and solve it using matrix techniques; it is less computationally intense to proceed as follows.

Let r, s, t, u be the given numbers and let a, b, c, d be the unknown numbers. Let v be the sum of the given numbers, which we can compute up front, and let e be the sum of the unknown numbers. Without loss of generality, assign

$$a = e - r, b = e - s, c = e - t, d = e - u.$$

Adding these equations gives $e = 4e - v$, so $3e = v$. So, $e = \frac{v}{3}$. From this, produce a, b, c, d.

For example, if $r = 22$, $s = 24$, $t = 27$, and $u = 20$, we have $v = 93$, so $e = 31$. Thus

$$a = 9, b = 7, c = 4, d = 11.$$

(c) Let $\theta = \angle CAD$, so that $2\theta = \angle CAB$. We want $\tan\theta = \frac{3}{4}$. Then

$$\tan 2\theta = \frac{2\tan\theta}{1 - \tan^2\theta} = \frac{24}{7}.$$

Let $CD = 3x$ so that $CA = 4x$. Let $y = CB$. Then

$$\tan 2\theta = \frac{y}{4x} = \frac{24}{7}.$$

We see that x must be a multiple of 7; trying $x = 7$, we have $y = 96$, and

$$(AB)^2 = 28^2 + 96^2 = 4^2(7^2 + 24^2) = 4^2(25^2).$$

Thus $x = 7$ produces an integer hypotenuse, and

$$AB = 100, AD = 35, AC = 28, BD = 75, DC = 21.$$

□

Problem 3. Using straightedge and compass, construct an angle of 54°. Describe each step, discussing why your construction is effective.

Solution. Construct a 72° angle as per previous instructions. The supplementary angle is 108°. Bisect this to obtain 54°. □

Definition 6. Let $m, n \in \mathbb{Z}$. The *least common multiple* of m and n is a positive integer $l \in \mathbb{Z}$ such that

(a) $m \mid l$ and $n \mid l$;
(b) $m \mid k$ and $n \mid k$ implies $l \mid k$.

Definition 7. Let $n \in \mathbb{Z}$ with $n \geq 2$. Set $\mathbb{Z}_n = \{r \in \mathbb{Z} \mid 0 \leq r < n\}$. Define a function

$$\rho_n : \mathbb{Z} \to \mathbb{Z}_n \quad \text{by} \quad \rho_n(a) = \text{ the remainder when } a \text{ is divided by } n.$$

We call ρ the *residue map.*

Definition 8. Let $m, n \in \mathbb{Z}$ with $m \geq 2$, $n \geq 2$. Define a function

$$\sigma_{m,n} : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n \quad \text{by} \quad \sigma_{m,n}(a) = (\rho_m(a), \rho_n(a)).$$

We call σ the *joint residue map.*

Problem 4. Let $m, n \in \mathbb{Z}$ with $m \geq 2$ and $n \geq 2$. Let $d = \gcd(m, n)$ and $l = \operatorname{lcm}(m, n)$.

(a) Show that if $d = 1$, then $\sigma_{m,n}$ is bijective.
(b) Show that if $a \equiv b \pmod{l}$, then $\sigma_{m,n}(a) = \sigma_{m,n}(b)$.

Proof. Fix m and n and let $\sigma = \sigma_{m,n}$. We note that $|\mathbb{Z}_{mn}| = mn = |\mathbb{Z}_m \times \mathbb{Z}_n|$. By a previous theorem, $\rho_n(a) = \rho_n(b)$ if and only if $a \equiv b \pmod{n}$.

(a) Let $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_n$. Since $\gcd(m, n) = 1$, the Chinese Remainder Theorem tells us that there exists $c \in \mathbb{Z}$ such that $a \equiv c \pmod{m}$ and $b \equiv c \pmod{n}$, that is, $\rho_m(c) = a$ and $\rho_n(c) = b$. Moreover, this c may be selected so that $0 \leq c < mn$; select c from $\mathbb{Z}_{mn}$. Then $\sigma(c) = (a, b)$, and σ is surjective. A surjective function between finite sets of the same cardinality is necessarily injective, so σ is bijective.

(b) Suppose k is a common multiple of m and n. Then there exist $x, y \in \mathbb{Z}$ such that $mx = k$ and $ny = k$.

We assume that $a \equiv bg \pmod{k}$, so $k \mid a - b$, and $a - b = kz$ for some $z \in \mathbb{Z}$. Thus $a - b = mxz$ and $a - b = nyz$. Thus $m \mid a - b$ and $n \mid a - b$. Therefore $a \equiv b \pmod{m}$ and $a \equiv b \pmod{n}$. □

Problem 2. Compute the volume of a regular icosahedron inscribed in a sphere of radius 1.

Solution. We proceed as follows.

- **(a)** Find an icosahedron inscribed in a sphere.
- **(b)** Find the area A of one face.
- **(c)** Find the length h of the apothem (the distance from the center of the sphere to the centroid of a face).
- **(d)** The volume of the tetrahedron whose base is a face and whose apex is the center of the sphere is $\frac{1}{3}Ah$. There are 20 faces, so the volume I of the entire icosahedron is $\frac{20Ah}{3}$.
- **(e)** Find the radius r of the sphere.
- **(f)** Find the volume V of the icosahedron inscribed in a unit sphere, which is $\frac{20Ah}{3r^3}$.

(a) *Find an icosahedron.* Let $\phi = \frac{1+\sqrt{5}}{2}$. Then $\phi^2 = \phi - 1$.

The twelve points $(\pm\phi, \pm 1, 0)$, $(\pm 1, 0, \pm\phi)$, $(0, \pm\phi, \pm 1)$, form the vertices of a regular icosahedron in $\mathbb{R}^3$.

(b) *Find the area A of one face.* Consider the face with vertices $(\phi, \pm 1, 0)$ and $(1, 0, \phi)$. The area of an equilateral triangle with edge length e is

$$A = \frac{1}{2}e(e\sin 60^\circ) = \frac{e^2\sqrt{3}}{4}.$$

The length of one side is distance between the first to vertices, which is

$$e = \sqrt{(\phi - \phi)^2 + (1 - (-1))^2} = \sqrt{4} = 2.$$

Thus $A = \sqrt{3}$.

(c) *Find the length h of the apothem.* The center of the sphere is the origin. The centroid is the average of the coordinates of the vertices, which is $(\frac{2\phi+1}{3}, 0, \frac{\phi}{3})$. The apothem is

$$h = \frac{1}{3}\sqrt{(2\phi + 1)^2 + \phi^2}.$$

We note that

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

Thus

$$(2\phi + 1)^2 + \phi^2 = 5\phi^2 + 4\phi + 1 = 3\phi(2\phi + 1) = 3\phi^4.$$

Therefore

$$h = \frac{1}{3}\sqrt{3\phi^4} = \frac{\phi^2}{\sqrt{3}}.$$

(d) *Find the volume of the tetrahedron.* We have

$$I = \frac{20Ah}{3} = \frac{20(\sqrt{3})(\phi^2/\sqrt{3})}{3} = \frac{20\phi^2}{3}.$$

(e) *Find the radius of the sphere.* This is the distance from the origin to a vertex, say $(1, 0, \phi)$. We have

$$r = \sqrt{\phi^2 + 1}.$$

Thus

$$V = \frac{20\phi^2}{3\sqrt{(\phi^2 + 1)^3}}.$$

□

Problem 3. Consider the elliptic curve given by the equation

$$y^2 = x^3 - 12x + 25.$$

Find as many rational points on this curve as you can, including all rational points that lie on a horizontal tangent. Justify your answer.

Solution. An *elliptic curve* is the locus to an equation of the form $y^2 = f(x)$, where $f(x)$ is a cubic polynomial. Elliptic curves play a critical role in advanced arithmetic geometry.

Let $f(x) = x^3 - ax + b$; we investigate methods to find critical points on the curve $C : y^2 = f(x)$.

If we find one rational point on C, then we can use it to reduce our (difficult) cubic equation to a (tractable) quadratic equation; however, we have no guarantee that this quadratic will yield rational results. On the other hand, if we have a double rational zero or two rational zeros of a cubic equation, we can reduce the quadratic to a linear equation, whose solution will necessarily be rational. Is is the tactic employed by Diophantus.

Let (p, q) be a rational point on C. Implicit differentiation gives the slope of the tangent line at this point to be

$$m = \frac{3p^2 - a}{2q}.$$

The line through this point is

$$L : y = mx + (q - mp).$$

We intersect the line L with the curve C.

Thus, for points on this line, we have $y^2 = m^2x^2 + 2m(q-mp)x + (q-mp)^2$. If $g(x) = m^2x^2 + 2m(q - mp) + (q - mp)^2$, then $f'(p) - g'(p) = 0$, so $f - g$ has a horizontal tangent at $x = p$; therefore, p is a double zero of $f - g$. In other words, L intersects C is at most one point other than (p, q).

Let $h(x) = f(x)-g(x) = x^3-m^2x^2+(2m^2p-2mq-a)+(b-(q-mp)^2)$. We divide $h(x)$ by $(x - p)^2$ (using synthetic division, we divide by p twice) and find that the quotient is $x+2p-m^2$. Therefore, $x = m^2-2p$ is another zero of $h(x)$; it is the x-coordinate of the other intersection point of L and C. The y-coordinate is obtained by plugging x into the line L, and we get $y = m(m^2-2p)+(q-mp)$. Thus, we have found another rational point:

By the tangent method: $m = \dfrac{3p^2 - a}{2q}$ giving $(m^2 - 2p, m^3 - 3mp + q)$.

Let (p_1, q_1) and (p_2, q_2) be rational points on C. Let

$$m = \frac{q_2 - q_1}{p_2 - p_1}.$$

Let $L : y = mx + (q_1 - mp_1)$. Again intersect L with C to construct h as above, and factor out $(x-p_1)$ and $(x-p_2)$. You will find that $x = m^2-p_1-p_2$ is the x-coordinate of the third point of intersection, and the corresponding y-coordinate is $m(x - p_1) + q_1$.

By the secant method: $m = \dfrac{q_2 - q_1}{p_2 - p_1}$ giving $(m^2-p_1-p_2, m^3-2mp_1-p_2+q_1)$.

Of course, it is fruitless to attempt to use the secant method on a pair of points where one has been derived from the other from the tangent method.

Now having developed these formulae, it was relatively easy to write a computer program to guess easy integer solutions and search for additional rational points using the tangent and secant method. The program first searches for solutions for x between -9 and 9, then builds up a list of all points found from these whose numerator and denominator have absolute value less than 10 million. □

Here is the source listing for the program to find rational points on $y^2 = x^3 - ax + b$.

```
// Find rational points on elliptic curve y^2 = x^3 - ax + b

#include <stdio.h>
#include <math.h>
#include "Rational.h"

// Find rational points on elliptic curve y^2 = x^3 - ax + b

Rational points[100][2];
int pointc=1;

Integer abs(Integer p)
{ if (p<0) return -p;
  return p; }

int find(Rational p)
{ int k=1;
  while (k<pointc)
  { if (points[k][0] == p) return k;
    k++; }
  return 0; }

int check(Rational p)
{ if (abs(p.Getm())>10000000 || abs(p.Getn())>10000000) return 1;
  return 0; }

int put(Rational p,Rational q)
{ if (pointc>98) return 1;
  if (find(p)) return 2;
  if (check(p)) return 3;
  if (check(q)) return 4;
  points[pointc][0] = p;
  points[pointc][1] = q;
  pointc++;
  printf("(%s,%s)\n",p.String(),q.String());
  return 0; }

int tangent(Rational a,Rational b,Rational p,Rational q)
{ Rational m,x,y;
  m = (3*p*p - a)/(2*q);
  x = m*m - 2*p;
  y = m*(x-p)+q;
  if (put(x,y)) return 0;
  return 1; }

int secant(Rational a,Rational b,Rational p1,Rational q1,Rational p2,Rational q2)
```

```
{ Rational m,x,y;
  m = (q2-q1)/(p2-p1);
  x = m*m - p1 - p2;
  y = m*(x-p1)+q1;
  if (put(x,y)) return 0;
  return 1; }

int search(Rational a,Rational b)
{ int i=0,j=0,k=0;
  Rational p1,q1,p2,q2,p,q;
  for (i=1; i<pointc-1; i++)
  { for (j=i+1; j<pointc; j++)
    { p1 = points[i][0];
      q1 = points[i][1];
      p2 = points[j][0];
      q2 = points[j][1];
      k += secant(a,b,p1,q1,p2,q2); } }
  for (i=1; i<pointc; i++)
  { p = points[i][0];
    q = points[i][1];
    k+= tangent(a,b,p,q); }
  return k; }

void elliptic(Integer a,Integer b)
{ Integer x,y,z;
  for (x=-9; x<=9; x++)
  { z = x*x*x - a*x + b;
    y = squr(z);
    if (y*y != z) continue;
    put(x,y); }
  while (search(a,b)); }

int main(int argc, char* argv[])
{ elliptic(12,25);
return 0; }
```

The output of the program is a list of the rational points it found.

```
(-4,3)
(-1,6)
(0,5)
(2,3)
(3,4)
(6,13)
(8,21)
(17/4,57/8)
(50/49,1275/343)
(-7/4,51/8)
(-38/9,17/27)
(-157/49,1896/343)
(152/121,4593/1331)
(-26/9,161/27)
(14/25,537/125)
(116/49,1077/343)
(1911/361,71878/6859)
(1529/5776,2051571/438976)
(-159/64,3217/512)
(36/25,409/125)
(-3592/1681,-440691/68921)
(-13192/3721,1088151/226981)
(-3382/961,-144861/29791)
(-663/2116,-521717/97336)
(16409/4624,1636989/314432)
(8018/1681,601941/68921)
(208/9,2969/27)
(2922/169,155137/2197)
(7728/2809,532831/148877)
(41/16,213/64)
(3014/3721,902559/226981)
(275,-4560)
(-1954/3025,948219/166375)
(122,1347)
(276/169,6863/2197)
(10225/324,-1028105/5832)
(44,291)
(14,51)
(1271/361,35238/6859)
(1124/289,29949/4913)
(7952/841,-670983/24389)
```

Problem 4. (Fibonacci)
Recall that the Fibonacci sequence (F_n) is defined by $F_1 = 1$, $F_2 = 1$, and $F_{n+2} = F_n + F_{n+1}$, and that $\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \phi$, where $\phi = \frac{1+\sqrt{5}}{2}$.

Let $b \in \mathbb{R}$ with $b \geq 1$ and define a sequence (G_n) by $G_1 = 1$, $G_2 = 1$, and $G_{n+2} = G_n + bG_{n+1}$.

Let $c \in \mathbb{R}$ with $c \geq \phi$. Find b such that $\lim_{n\to\infty} \frac{G_{n+1}}{G_n} = c$.

Solution. Let $c_n = \frac{G_{n+1}}{G_n}$. Then

$$c_{n+1} = \frac{G_{n+2}}{G_{n+1}} = \frac{bG_{n+1} + G_n}{G_{n+1}} = b + \frac{1}{c_n}.$$

Now (c_n) is a Cauchy sequence, so it converges; let $L = \lim c_n$. Since $c_n > 0$ for all n, $L \geq 0$. Then $L = b + \frac{1}{L}$, so

$$L^2 - bL - 1 - 0.$$

Thus

$$L = \frac{b + \sqrt{b^2+4}}{2}.$$

If $L = c$, then $2c = b + \sqrt{b^2+4}$, so $(2c-b)^2 = b^2 + 4$, so $4c^2 - 4bc + b^2 = b^2 + 4$, so

$$\boxed{b = \frac{c^2 - 1}{c}.}$$

□

Problem 5. (Tartaglia)
Recall that Tartaglia viewed the cube x^3 as $(t-u)^3$ to find solutions to cubic equations.

Let $f(x) = x^3 + 3x^2 + 6x - 8$. Find the real zero of f using Tartaglia's cube plus cosa method.

Solution. First we depress the cubic: let $y = x + 1$ so that $x = y - 1$; then

$$\begin{aligned} f(x) &= f(y-1) \\ &= (y-1)^3 + 3(y-1)^2 + 6(y-1) - 8 \\ &= y^3 - 3y^2 + 3y - 1 + 3y^2 - 6y + 3 + 6y - 6 - 8 \\ &= y^3 + 3y - 12. \end{aligned}$$

We now solve $y^3 + 3y = 12$. Set $3tu = 3$ and $t^3 - u^3 = 12$, so that $u = \frac{1}{t}$, and $t^3 - \frac{1}{t^3} = 12$. Thus

$$t^6 - 12t^3 - 1 = 0.$$

By the quadratic formula,

$$t^3 = \frac{12 + \sqrt{144+4}}{2} = 6 + \sqrt{37}.$$

Now $u^3 = t^3 - 12 = -6 + \sqrt{37}$. Thus

$$y = t - u = \sqrt[3]{6+\sqrt{37}} + \sqrt[3]{6-\sqrt{37}}.$$

Finally,

$$\boxed{x = \sqrt[3]{6+\sqrt{37}} + \sqrt[3]{6-\sqrt{37}} - 1.}$$

□

Problem 6. (Descartes)
Recall that Descartes used the concept of expanding circles and the ability to compute the number of real solutions to quadratic equations to find tangents.

Find the distance between the curve $x = y^2$ and the point $(3, 0)$ using Descartes' discriminant method.

Solution. A circle of radius r centered at $(3, 0)$ has equation $(x-3)^2 + y^2 = r^2$. The shortest distance to the curve is the radius of a tangential circle, which occurs when then circle intersects the curve in exactly one point.

Intersecting the curve and the circle gives $(x-3)^2 + x = r^2$, so $x^2 - 5x + (9 - r^2) = 0$, so $x = \frac{5 \pm \sqrt{25 - 4(9-r^2)}}{2}$. This has exactly one solution when $25 = 4(9 - r^2)$, or $r^2 = 9 - \frac{25}{4} = \frac{11}{4}$. Thus the distance is

$$\boxed{r = \frac{\sqrt{11}}{2}.}$$

□

Problem 7. (Napier)
Recall that Napier desired to find a function to convert multiplication into addition. We may use techniques of Calculus unavailable to him to see that he had very little choice. The modern definition is

$$\log x = \int_1^x \frac{dt}{t} \quad \text{and} \quad \log_b(x) = \frac{\log x}{\log b}.$$

Let $f : (0, \infty) \to \mathbb{R}$ be a differentiable function which is not constantly zero and satisfies

$$f(ab) = f(a) + f(b) \quad \text{for all } a, b \in (0, \infty).$$

Show that there exists $b \in \mathbb{R}$ such that $f(x) = \log_b(x)$.

Solution. First, note that $f(1) = f(1 \cdot 1) = f(1) + f(1)$; thus $f(1) = 0$.

Fix $t \in (0, \infty)$; we have $f(tx) = f(t) + f(x)$. Differentiating with respect to x gives $tf'(tx) = f'(x)$. In particular, if $x = 1$, we have $tf'(t) = f'(1)$, so $f'(t) = \frac{f'(1)}{t}$. This is true for all $t \in \mathbb{R}$, so

$$\int_1^x f'(t)\, dt = \int_1^x \frac{f'(1)}{t}\, dt.$$

By the Fundamental Theorem of Calculus,

$$f(x) - f(1) = f'(1) \int_1^x \frac{dt}{t} = f'(1) \log x.$$

Since $f(1) = 0$, $f(x) = f'(1) \log x$.

Suppose $f'(1) = 0$; then $tf'(t) = 0$, so $f'(t) = 0$ for all $t \in (0, \infty)$, so f is constant. But $f(1) = 0$, so $f(x) = 0$; this contradicts that f is nonzero. Thus $f'(1) \neq 0$.

Let $b = e^{\frac{1}{f'(1)}}$. Then $f'(1) = \frac{1}{\log b}$, and $f(x) = \frac{\log x}{\log b}$; that is,

$$\boxed{f(x) = \log_b x \quad \text{where } b = e^{\frac{1}{f'(1)}}.}$$

□

APPENDIX F

Additional Material

1. Plimpton Tablet

n	b	c	s	a	u	v	t
1	119	169	1.983403	120	12	5	44.760308
2	3367	4825	1.949159	3456	64	27	44.252707
3	4601	6649	1.918802	4800	75	32	43.787383
4	12709	18541	1.886248	13500	125	54	43.271348
5	65	97	1.815008	72	9	4	42.075058
6	319	481	1.785193	360	20	9	41.544544
7	2291	3541	1.719984	2700	54	25	40.315256
8	799	1249	1.692709	960	32	15	39.770364
9	481	769	1.642669	600	25	12	38.718021
10	4961	8161	1.586123	6480	81	40	37.437210
11	3	5	1.562500	4	2	1	36.869929
12	1679	2929	1.489417	2400	48	25	34.976024
13	161	289	1.450017	240	15	8	33.855055
14	1771	3229	1.430239	2700	50	27	33.261936
15	56	106	1.387160	90	9	5	31.890819

2. Euclid's Definitions

Definitions

Definition 1.
A point is that which has no part.

Definition 2.
A line is breadthless length.

Definition 3.
The ends of a line are points.

Definition 4.
A straight line is a line which lies evenly with the points on itself.

Definition 5.
A surface is that which has length and breadth only.

Definition 6.
The edges of a surface are lines.

Definition 7.
A plane surface is a surface which lies evenly with the straight lines on itself.

Definition 8.
A plane angle is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.

Definition 9.
And when the lines containing the angle are straight, the angle is called rectilinear.

Definition 10.
When a straight line standing on a straight line makes the adjacent angles equal to one another, each of the equal angles is right, and the straight line standing on the other is called a perpendicular to that on which it stands.

Definition 11.
An obtuse angle is an angle greater than a right angle.

Definition 12.
An acute angle is an angle less than a right angle.

Definition 13.
A boundary is that which is an extremity of anything.

Definition 14.
A figure is that which is contained by any boundary or boundaries.

Definition 15.
A circle is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure equal one another.

Definition 16.
And the point is called the center of the circle.

Definition 17.
A diameter of the circle is any straight line drawn through the center and terminated in both directions by the circumference of the circle, and such a straight line also

bisects the circle.

Definition 18.
A semicircle is the figure contained by the diameter and the circumference cut off by it. And the center of the semicircle is the same as that of the circle.

Definition 19.
Rectilinear figures are those which are contained by straight lines, trilateral figures being those contained by three, quadrilateral those contained by four, and multilateral those contained by more than four straight lines.

Definition 20.
Of trilateral figures, an equilateral triangle is that which has its three sides equal, an isosceles triangle that which has two of its sides alone equal, and a scalene triangle that which has its three sides unequal.

Definition 21.
Further, of trilateral figures, a right-angled triangle is that which has a right angle, an obtuse-angled triangle that which has an obtuse angle, and an acute-angled triangle that which has its three angles acute.

Definition 22.
Of quadrilateral figures, a square is that which is both equilateral and right-angled; an oblong that which is right-angled but not equilateral; a rhombus that which is equilateral but not right-angled; and a rhomboid that which has its opposite sides and angles equal to one another but is neither equilateral nor right-angled. And let quadrilaterals other than these be called trapezia.

Definition 23
Parallel straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction.

Postulates

Postulate 1.
To draw a straight line from any point to any point.

Postulate 2.
To produce a finite straight line continuously in a straight line.

```
Postulate 3.
    To describe a circle with any center and radius.

Postulate 4.
    That all right angles equal one another.

Postulate 5.
    That, if a straight line falling on two straight lines makes
    the interior angles on the same side less than two right
    angles, the two straight lines, if produced indefinitely,
    meet on that side on which are the angles less than the two
    right angles.

Common Notions

Common notion 1.
    Things which equal the same thing also equal one another.

Common notion 2.
    If equals are added to equals, then the wholes are equal.

Common notion 3.
    If equals are subtracted from equals, then the remainders are
    equal.

Common notion 4.
    Things which coincide with one another equal one another.

Common notion 5.
    The whole is greater than the part.

Propositions

Proposition 1.
    To construct an equilateral triangle on a given finite
    straight line.

Proposition 2.
    To place a straight line equal to a given straight line
    with one end at a given point.

Proposition 3.
    To cut off from the greater of two given unequal straight
    lines a straight line equal to the less.

Proposition 4.
    If two triangles have two sides equal to two sides
    respectively, and have the angles contained by the equal
```

straight lines equal, then they also have the base equal to the base, the triangle equals the triangle, and the remaining angles equal the remaining angles respectively, namely those opposite the equal sides.

Proposition 5.
In isosceles triangles the angles at the base equal one another, and, if the equal straight lines are produced further, then the angles under the base equal one another.

Proposition 6.
If in a triangle two angles equal one another, then the sides opposite the equal angles also equal one another.

Proposition 7.
Given two straight lines constructed from the ends of a straight line and meeting in a point, there cannot be constructed from the ends of the same straight line, and on the same side of it, two other straight lines meeting in another point and equal to the former two respectively, namely each equal to that from the same end.

Proposition 8.
If two triangles have the two sides equal to two sides respectively, and also have the base equal to the base, then they also have the angles equal which are contained by the equal straight lines.

Proposition 9.
To bisect a given rectilinear angle.

Proposition 10.
To bisect a given finite straight line.

Proposition 11.
To draw a straight line at right angles to a given straight line from a given point on it.

Proposition 12.
To draw a straight line perpendicular to a given infinite straight line from a given point not on it.

Proposition 13.
If a straight line stands on a straight line, then it makes either two right angles or angles whose sum equals two right angles.

Proposition 14.

If with any straight line, and at a point on it, two straight lines not lying on the same side make the sum of the adjacent angles equal to two right angles, then the two straight lines are in a straight line with one another.

Proposition 15.
If two straight lines cut one another, then they make the vertical angles equal to one another.

Corollary. If two straight lines cut one another, then they will make the angles at the point of section equal to four right angles.

Proposition 16.
In any triangle, if one of the sides is produced, then the exterior angle is greater than either of the interior and opposite angles.

Proposition 17.
In any triangle the sum of any two angles is less than two right angles.

Proposition 18.
In any triangle the angle opposite the greater side is greater.

Proposition 19.
In any triangle the side opposite the greater angle is greater.

Proposition 20.
In any triangle the sum of any two sides is greater than the remaining one.

Proposition 21.
If from the ends of one of the sides of a triangle two straight lines are constructed meeting within the triangle, then the sum of the straight lines so constructed is less than the sum of the remaining two sides of the triangle, but the constructed straight lines contain a greater angle than the angle contained by the remaining two sides.

Proposition 22.
To construct a triangle out of three straight lines which equal three given straight lines: thus it is necessary that the sum of any two of the straight lines should be greater than the remaining one.

Proposition 23.
To construct a rectilinear angle equal to a given rectilinear angle on a given straight line and at a point on it.

Proposition 24.
If two triangles have two sides equal to two sides respectively, but have one of the angles contained by the equal straight lines greater than the other, then they also have the base greater than the base.

Proposition 25.
If two triangles have two sides equal to two sides respectively, but have the base greater than the base, then they also have the one of the angles contained by the equal straight lines greater than the other.

Proposition 26.
If two triangles have two angles equal to two angles respectively, and one side equal to one side, namely, either the side adjoining the equal angles, or that opposite one of the equal angles, then the remaining sides equal the remaining sides and the remaining angle equals the remaining angle.

Proposition 27.
If a straight line falling on two straight lines makes the alternate angles equal to one another, then the straight lines are parallel to one another.

Proposition 28.
If a straight line falling on two straight lines makes the exterior angle equal to the interior and opposite angle on the same side, or the sum of the interior angles on the same side equal to two right angles, then the straight lines are parallel to one another.

Proposition 29.
A straight line falling on parallel straight lines makes the alternate angles equal to one another, the exterior angle equal to the interior and opposite angle, and the sum of the interior angles on the same side equal to two right angles.

Proposition 30.
Straight lines parallel to the same straight line are also parallel to one another.

Proposition 31.
To draw a straight line through a given point parallel to a given straight line.

Proposition 32.
In any triangle, if one of the sides is produced, then the exterior angle equals the sum of the two interior and opposite angles, and the sum of the three interior angles of the triangle equals two right angles.

Proposition 33.
Straight lines which join the ends of equal and parallel straight lines in the same directions are themselves equal and parallel.

Proposition 34.
In parallelogrammic areas the opposite sides and angles equal one another, and the diameter bisects the areas.

Proposition 35.
Parallelograms which are on the same base and in the same parallels equal one another.

Proposition 36.
Parallelograms which are on equal bases and in the same parallels equal one another.

Proposition 37.
Triangles which are on the same base and in the same parallels equal one another.

Proposition 38.
Triangles which are on equal bases and in the same parallels equal one another.

Proposition 39.
Equal triangles which are on the same base and on the same side are also in the same parallels.

Proposition 40.
Equal triangles which are on equal bases and on the same side are also in the same parallels.

Proposition 41.
If a parallelogram has the same base with a triangle and is in the same parallels, then the parallelogram is double the triangle.

Proposition 42.
To construct a parallelogram equal to a given triangle in a given rectilinear angle.

```
Proposition 43.
    In any parallelogram the complements of the parallelograms
    about the diameter equal one another.

Proposition 44.
    To a given straight line in a given rectilinear angle, to
    apply a parallelogram equal to a given triangle.

Proposition 45.
    To construct a parallelogram equal to a given rectilinear
    figure in a given rectilinear angle.

Proposition 46.
    To describe a square on a given straight line.

Proposition 47.
    In right-angled triangles the square on the side opposite the
    right angle equals the sum of the squares on the sides
    containing the right angle.

Proposition 48.
    If in a triangle the square on one of the sides equals the sum
    of the squares on the remaining two sides of the triangle,
    then the angle contained by the remaining two sides of the
    triangle is right.
```

3. Eudoxus Theory of Proportion Definition

Eudoxus invented the "theory of proportion"; this is the definition for Euclid's Elements:

Def. 5. Magnitudes are said to be in the same ratio, the first to the second and the third to the fourth, when, if any equimultiples whatever are taken of the first and third, and any equimultiples whatever of the second and fourth, the former equimultiples alike exceed, are alike equal to, or alike fall short of, the latter equimultiples respectively taken in corresponding order.

http://www.mathcs.clarku.edu/ djoyce/java/elements/bookV/defV5.html

Bibliography

[Be37] Bell, E.T., *Men of Mathematics*, Simon and Schuster (1937) recopywrighted (1965)

[De02] Devlin, Keith, *The Millennium Problems*, Basic Books, Perseus Books Group (2002)

[Du90] Dunham, William, *Journey Through Genius*, Penguin Books (1990)

[Ev90] Eves, Howard, *An Introduction to the History of Mathematics*, 6^{th} edition, Brooks/Cole, Thompson (1990)

[Ha78] Hadlock, Charles Robert, *Field Theory and its Classical Problems*, The Carus Mathematical Monographs, Mathematical Association of America (1978)

[Li02] Livio, Mario *The Golden Ratio*, Broadway Books (2002)

[Sm29] Smith, David Eugene, *A Source Book in Mathematics*, McGraw-Hill (1929) reprinted by Dover (1959)