Undergraduate Texts in Mathematics

Springer Science+Business Media, LLC

Undergraduate Texts in Mathematics

Anglin: Mathematics: A Concise History and Philosophy.
Readings in Mathematics.
Anglin/Lambek: The Heritage of Thales.
Readings in Mathematics.
Apostol: Introduction to Analytic Number Theory. Second edition.
Armstrong: Basic Topology.
Armstrong: Groups and Symmetry.
Axler: Linear Algebra Done Right.
Beardon: Limits: A New Approach to Real Analysis.
Bak/Newman: Complex Analysis. Second edition.
Banchoff/Wermer: Linear Algebra Through Geometry. Second edition.
Berberian: A First Course in Real Analysis.
Brémaud: An Introduction to Probabilistic Modeling.
Bressoud: Factorization and Primality Testing.
Bressoud: Second Year Calculus.
Readings in Mathematics.
Brickman: Mathematical Introduction to Linear Programming and Game Theory.
Browder: Mathematical Analysis: An Introduction.
Buskes/van Rooij: Topological Spaces: From Distance to Neighborhood.
Cederberg: A Course in Modern Geometries.
Childs: A Concrete Introduction to Higher Algebra. Second edition.
Chung: Elementary Probability Theory with Stochastic Processes. Third edition.
Cox/Little/O'Shea: Ideals, Varieties, and Algorithms. Second edition.
Croom: Basic Concepts of Algebraic Topology.
Curtis: Linear Algebra: An Introductory Approach. Fourth edition.
Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.
Dixmier: General Topology.
Driver: Why Math?
Ebbinghaus/Flum/Thomas: Mathematical Logic. Second edition.
Edgar: Measure, Topology, and Fractal Geometry.
Elaydi: Introduction to Difference Equations.
Exner: An Accompaniment to Higher Mathematics.
Fine/Rosenberger: The Fundamental Theory of Algebra.
Fischer: Intermediate Real Analysis.
Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.
Fleming: Functions of Several Variables. Second edition.
Foulds: Combinatorial Optimization for Undergraduates.
Foulds: Optimization Techniques: An Introduction.
Franklin: Methods of Mathematical Economics.
Gordon: Discrete Probability.
Hairer/Wanner: Analysis by Its History.
Readings in Mathematics.
Halmos: Finite-Dimensional Vector Spaces. Second edition.
Halmos: Naive Set Theory.
Hämmerlin/Hoffmann: Numerical Mathematics.
Readings in Mathematics.
Hijab: Introduction to Calculus and Classical Analysis.
Hilton/Holton/Pedersen: Mathematical Reflections: In a Room with Many Mirrors.
Iooss/Joseph: Elementary Stability and Bifurcation Theory. Second edition.
Isaac: The Pleasures of Probability.
Readings in Mathematics.

(continued after index)

George E. Martin

Geometric Constructions

With 112 figures

Springer

George E. Martin
Department of Mathematics and Statistics
State University of New York at Albany
Albany, NY 12222
USA

Mathematics Subject Classification (1991): 51-01

Library of Congress Cataloging-in-Publication Data
Martin, George E., 1932–
Geometric constructions / George E. Martin.
p. cm. – (Undergraduate texts in mathematics)
Includes bibliographical references and index.
ISBN 978-1-4612-6845-1 ISBN 978-1-4612-0629-3 (eBook)
DOI 10.1007/978-1-4612-0629-3
1. Geometrical constructions. I. Title. II. Series.
QA464.M285 1997
516—dc21 97-22885
CIP

Printed on acid-free paper.

Production managed by Victoria Evarretta; manufacturing supervised by Joe Quatela.
Photocomposed pages prepared from author's LaTex files.

9 8 7 6 5 4 3 2 1

SPIN 10557774

To Margaret

Books by the Author

The Foundations of Geometry
and the Non-Euclidean Plane

Transformation Geometry,
An Introduction to Symmetry

Polyominmoes,
A Guide to Puzzles and Problems in Tiling

Geometric Constructions

Contents

Preface

> Books are to be called for and supplied on the assumption that the process of reading is not half-sleep, but in the highest sense an exercise, a gymnastic struggle; that the reader is to do something for himself.
>
> WALT WHITMAN

The old games are the best games. One of the oldest is geometric constructions. As specified by Plato, the game is played with a ruler and a compass, where the ruler can be used only to draw the line through two given points and the compass can be used only to draw the circle with a given center and through a given point. Skilled players of the game sometimes give themselves a handicap, such as restricting the compass to a fixed opening. A more severe restriction is to use only the ruler, after drawing exactly one circle (Chapter 6). On the other hand, a master player of Plato's game need not use the ruler at all (Chapter 3). Some prefer to play the game of geometric constructions with other tools, even toothpicks (Chapter 8). The most famous of the other construction tools is the marked ruler, which is simply a ruler with two marks on its edge (Chapter 9). We can do more constructions with only the marked ruler than with the ruler and compass. For example, we will prove that angle trisection is generally impossible with only the ruler and compass (Chapter 2), and we will see how to trisect any given angle with a marked ruler. The first chapter starts from scratch and reminds us of all the euclidean constructions from high school that we have forgotten or never seen. The last chapter covers geometric constructions by paperfolding.

Although many of our construction problems are inherited from antiquity, we take advantage of modern algebra and the resultant coordinate geometry to analyze and classify these problems. We necessarily encounter algebra in exploring the constructions. Various geometric construction tools are associated with various algebraic fields of numbers. This book is about these associations. Some readers will find this theoretical association a fascinating end in itself. Some will be stimulated to seek out elegant means of accomplishing those constructions that the theory proves exist and will know to avoid those proposed constructions that the theory proves do not exist. It is important to know what cannot be done in order to avoid wasting time in attempting impossible constructions. The reader of this book will not be among those few persons who turn up every year to proclaim they have "solved a construction problem that has stumped mathematicians for over two thousand years." The principal purposes for reading this book are to learn a little geometry and a little algebra and to enjoy the exercise.

Very little mathematical background is required of the reader. Abstract algebra, in general, and galois theory, in particular, are not prerequisite. Once the ideas introduced in the second chapter become familiar, the rest of the book follows smoothly. Even though the format is that of a textbook, there are so many hints and answers to be found in the lengthy section called The Back of the Book that the individual studying alone should have no problem testing comprehension against some of the exercises. A lozenge ◊ indicates that a given exercise has an entry in The Back of the Book.

By skipping over the optional Chapter 8 to get to the essential Chapter 9, an instructor can expect to cover the material in one semester. A new instructor should be warned that, although students will at first balk at the schemes that are introduced in the first chapter, the students will very quickly learn to use them and that the instructor's problem will be turning the schemes off when they are no longer appropriate.

If the figures in the text have a home-made look, it is partly because they have been made by an author learning to use *The Geometer's Sketchpad, Dynamic Geometry for the Macintosh*, published by Key Curriculum Press. The dynamic power of this software helped in making the figures and suggests a challenging follow-up seminar that attacks the question, What points can theoretically be constructed with this software? The task would be to consider the mathematical aspects of formulating a new chapter with the geometric construction tool motivated by *The Geometer's Sketchpad.*

The material has been class tested for many semesters with a master's level class for secondary teachers. The students in these classes have helped shape this book. The text jelled in the summer of 1984 with the then new Macintosh. Notes from that time show that we had class elections to determine the official definitions for the semester. The preliminary version of the text then carried the dedication FOR GILLYGALOOS EVERYWHERE. A residue of these classes can be seen in the somewhat unconventional Chapter 7, where there is a possibility of hands-on learning about mathematical structure.

I would like to thank the editors at Springer-Verlag for accepting *Geometric Constructions* for this distinguished series. There are three wonderful women at Springer-Verlag New York who have steered the text from manuscript to bound book. They are Ina Lindemann, Anne Fossella, and Victoria Evarretta. I also wish to thank Mademoiselle Claude Jacir, Documentaliste au Musée, l'Ordre de la Légion d'honneur, for providing information on Pierre Joseph Glotin. Finally, I am very much indebted to my friend and colleague Hugh Gordon, who made many helpful suggestions while teaching from preliminary versions of this book.

George E. Martin
martin@math.albany.edu

1

Euclidean Constructions

> There is much to be said in favor of a game you play alone ... the company is most congenial and perfectly matched in skill and intelligence, and there is no embarrassing sarcastic utterance should you make a stupid play. The game is particularly good if it is truly challenging and if it possesses manifold variety. ... The Greek geometers of antiquity devised such a game ...
>
> HOWARD EVES

Until recently, Euclid's name and the word *geometry* were synonymous. It was Euclid who first placed mathematics on an axiomatic basis. He did such a remarkable job of presenting much of the known mathematical results of his time in such an excellent format that almost all the mathematical works that preceded his were discarded. Euclid's principal work, *Elements*, has been the dominating text in mathematics for twenty-three centuries. It is only in this last half of our own century that Euclid is not the primary text used by beginning students. Yet, we know almost nothing about this person who wrote the world's most successful secular book. We suppose Euclid studied at Plato's Academy in Athens and was an early member of the famed Museum/Library at Alexandria. Alexandria became the most important city in the Western world after the death of Alexander the Great and remained so until Caesar's Rome dominated Cleopatra's City. Even

G. E. Martin, *Geometric Constructions*

then, while Rome was at its height, Alexandria remained the intellectual capital of the Empire. Alexandria was a major influence for a thousand years, from the time of Euclid in 300 BC until the fall of Alexandria to the Arabs in AD 641. Greek mathematics is mostly a product of the Golden Age of Greek Science and Mathematics, which was centered at Alexandria in the third century BC. Although located in what is now Egypt, the ancient city was Greek with the full name Alexandria-near-Egypt. The first of the city's rulers who could even speak to the Egyptians in their own language was Cleopatra, who died in 30 BC. Can there be any doubt that the very learned Cleopatra studied her geometry from Euclid's *Elements*?

Euclid's *Elements* is divided into thirteen Books, preceded by the Axioms and the Postulates. Although there has been a great deal written about the difference between these two types of fundamental assumptions, today we no longer debate about which should be which and use the words *axiom* and *postulate* interchangeably to denote an underlying assumption.

A *definition* is an abbreviation. Definitions may abbreviate mathematical concepts with symbols as well as with words, which are, after all, also symbols. For example, assuming we understand what it means to talk about the vertices of a triangle, the mathematical symbol $\triangle ABC$ is defined to be an abbreviation for "the triangle whose vertices are the points A,B,C." Some maintain that the principal art of creating mathematics is formulating the proper definitions. All mathematics students know that the first thing that must be mastered in any mathematics class is the definitions. Otherwise, the student fails to understand what is being discussed. As Socrates said, The beginning of wisdom is the definition of terms.

Each Book of Euclid's *Elements* contains a sequence of Propositions, which are of two types: the theorems and the problems. In general, a *theorem* is a statement that has a proof based on a given set of postulates and previously proved theorems. A *proof* is a convincing argument. A *problem* in Euclid asks that some new geometric entity be created from a given set. We call a solution to such a problem a *construction*. This construction is itself a theorem, requiring a proof and having the form of a recipe: If you do this, this, and this, then you will get that. Such a mathematical recipe is called an algorithm. So a construction is the special type of theorem that is also an algorithm. (We hesitatingly offer the analogy: *Problem*: Make a pudding; *Construction*: Recipe; *Proof*: Eating.)

It may be worthwhile to reread the preceding paragraph. For a simple example of this important Problem–Construction–Proof sequence, we can take Euclid's first proposition: *Problem: Given two points A and B, construct a point C such that $\triangle ABC$ is equilateral.* After introducing the notation P_Q for the circle with center P that passes through point Q, the appropriate theorem is easily stated: *Construction: If A and B are two points and if C is one of the points of intersection of the circles A_B and B_A, then $\triangle ABC$ is equilateral.* In particular, the construction not only states the existence of the point C required by the problem but also explicitly tells how to find C. The argument for this construction is short: *Proof: Since* $AC = AB$ *and* $BC = BA$ *because radii of the same circle are congruent, then* $AB = BC = CA$.

In addition to the Problem–Construction–Proof sequence, we should not overlook the fun of actually representing the theorem by creating an illustration that is carefully drawn with the geometric tools. Usually this drawing is also called a *construction* (The pudding?). So "construction" has two technical meanings. We will use the word for both meanings. Whether a particular occurrence of the word means the special type of theorem that is a geometric algorithm or means a drawing that illustrates such a theorem must be determined from the context. It is the combination of the construction (theorem) and the construction (drawing) that has provided so much pleasure to so many persons for so many centuries. You may feel that one is incomplete without the other.

The term *sketch* is reserved in this book for an informal, freehand representation of a formal construction drawing. Sketches are usually sufficient for our purposes. As a final observation about words and their meanings, we note that there will probably be no confusion with the general use of the word *figure* and its technical use, where the word denotes a set of points in the plane.

Whether a constructed drawing or a freehand sketch is used to illustrate a theorem, we have almost certainly been warned often that we should not argue from the figure. Some have even suggested that there should be no figures in geometry to avoid the temptation to this fault. However, figures not only help us keep track of complicated algorithms, but they are fun to see and fun to create. Geometry without figures is possible but not enjoyable. This is especially true when the geometry concerns the topic of geometric constructions.

We will use the notation in the adjacent box throughout.

A–B–C	means point B is between points A and C.
$\overleftrightarrow{AB}$	denotes the line through the two points A and B.
$\overrightarrow{AB}$	denotes the ray with vertex A that passes through B.
$\overline{AB}$	denotes the segment with endpoints A and B.
AB	denotes the distance from A to B.
A_B	denotes the circle through B with center A.
A_{BC}	denotes the circle with center A and radius BC.
$m\angle ABC$	denotes the degree measure of the angle $\angle ABC$.
$\triangle ABC$	denotes the triangle with vertices A,B,C.
ABC	denotes the area of $\triangle ABC$.
$\square ABCD$	denotes the quadrilateral with sides $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, $\overline{DA}$.
$ABCD$	denotes the area of $\square ABCD$.
$p = q$	means "p" and "q" are names for the same object.
$p \cong q$	is read "p is congruent to q."
$p \sim q$	is read "p is similar to q."

By "given a segment," we suppose the two endpoints of a segment are determined. In the absence of further explanation, we understand "given $\overleftrightarrow{AB}$" to mean only that the two points A and B are determined.

Almost everyone knows that "Q.E.D." signifies the end of an argument. The end of the proof of a theorem in Euclid is traditionally noted by this abbreviation for "quod erat demonstratum" (which was to be proved). However, not many know what "Q.E.F." signifies. This is short for "quod erat factorum" (which was to be constructed) and comes at the end of the proof of a construction that solves a problem. Of course, this notation is a tradition from the Latin translations and not the original Greek. Today, the end of a proof is generally denoted by a symbol like the one we will use: $\blacksquare$.

For easy reference, we are going to state in one location many of the construction problems from Euclid. These are selected from the thirteen Books that constitute the *Elements.* A Roman numeral is used to denote the Book in which a proposition can be found. For example, "VI.8" denotes the eighth proposition in Book VI of Euclid's *Elements.*

All the Construction Problems from Book I

Euclid I.1. Construct an equilateral triangle having a given segment as one side.

Euclid I.2. Construct a segment congruent to a given segment and with a given point as one endpoint.

Euclid I.3. Given $\overrightarrow{AB}$ and $\overline{CD}$, construct point E on $\overrightarrow{AB}$ such that $\overline{AE} \cong \overline{CD}$.

Euclid I.9. Construct the angle bisector of a given angle.

Euclid I.10. Construct the midpoint of a given segment.

Euclid I.11. Given $\overleftrightarrow{AB}$, construct the perpendicular to $\overleftrightarrow{AB}$ at A.

Euclid I.12. Given point C off $\overleftrightarrow{AB}$, construct the perpendicular to $\overleftrightarrow{AB}$ that passes through C.

Euclid I.22. Construct a triangle having sides respectively congruent to three given segments whose lengths are such that the sum of the lengths of any two is greater than the length of the third.

Euclid I.23. Given $\overline{AB}$ and $\angle CDE$, construct a point F such that $\angle FAB \cong \angle CDE$.

Euclid I.31. Through a given point, construct the parallel to a given line.

Euclid I.42. Construct a parallelogram having an angle congruent to a given angle and having the area of a given triangle.

Euclid I.44. Construct a parallelogram having a given segment as a side, having an angle congruent to a given angle, and having the area of a given triangle.

Euclid I.45. Construct a parallelogram having an angle congruent to a given angle and having the area of a given polygon.

Euclid I.46. Construct a square having a given segment as one side.

Construction Problems Selected from Books II, III, and VI

Euclid II.11. Given $\overline{AB}$, construct point X on $\overline{AB}$ such that $(AB)(BX) = (AX)^2$.

Euclid II.14. Construct a square with an area equal to that of a given polygon.

Euclid III.1. Given three noncollinear points, construct the center of the circle containing the three points.

Euclid III.17. Through a given point outside a given circle, construct a tangent to the circle.

Euclid VI.12. Construct a fourth proportional to three given segments.

Euclid VI.13. Construct a mean proportional to two given segments.

All the Propositions of Book IV

Euclid IV.1. In a given circle, inscribe a chord congruent to a given segment that is shorter than a diameter.

Euclid IV.2. In a given circle, inscribe a triangle equiangular with a given triangle.

Euclid IV.3. About a given circle, circumscribe a triangle equiangular with a given triangle.

Euclid IV.4. In a given triangle, inscribe a circle.

Euclid IV.5. About a given triangle, circumscribe a circle.

Euclid IV.6. In a given circle, inscribe a square.

Euclid IV.7. About a given circle, circumscribe a square.

Euclid IV.8. In a given square, inscribe a circle.

Euclid IV.9. About a given square, circumscribe a circle.

Euclid IV.10. Construct an isosceles triangle having base angles that are double the third angle.

Euclid IV.11. In a given circle, inscribe a regular pentagon.

Euclid IV.12. About a given circle, circumscribe a regular pentagon.

Euclid IV.13. In a given regular pentagon, inscribe a circle.

Euclid IV.14. About a given regular pentagon, circumscribe a circle.

Euclid IV.15. In a given circle, inscribe a regular hexagon.

Euclid IV.16. In a given circle, inscribe a regular pentadecagon.

What tools are available for these constructions? Although we will consider other possibilities in later chapters, even the answer "the ruler and the compass" may need some explaining. A *euclidean ruler* is used only to draw the line through any two given points. A physical model of the euclidean ruler has no marks on it and is sometimes called a *straightedge.* Of course, such a physical model is necessarily of finite length, unlike the ideal euclidean ruler. For us, the word *ruler* alone will always mean a euclidean ruler. Constructions using a marked ruler will be considered in Chapter 9. A *dividers* is the drafter's tool that accomplishes the construction for Euclid I.3; a dividers is used to "carry distance." Constructions using a dividers will be investigated in Chapter 5. The *modern compass*, the compass we buy in the school supply department (and which is completely adequate for any of our needs here), serves as a dividers as well as for the purpose of drawing circles. With a modern compass, we can draw a circle having a given center and having radius the length of a given segment. The *euclidean compass*, on the other hand, can be used only to draw the circle that passes through a given point and that has a given point as its center. Note that a euclidean compass has the peculiarity of collapsing when lifted and cannot be used as a dividers. For two reasons, the word *compass* alone will always mean a modern compass for us. First, it is difficult to imagine a physical model of a euclidean compass. The second and more important reason is that Euclid's constructions for I.1, I.2, and I.3 require only a ruler and a euclidean compass. Euclid shows in his first two propositions that the same circles can be constructed with a ruler and euclidean compass as can be constructed with a ruler and modern compass, although it may take more operations using the euclidean compass. The third proposition then shows that the euclidean compass also has the power to "carry distance."

Euclid I.2. Construct a segment congruent to a given segment and with a given point as one endpoint.

Euclid I.3. Given $\overrightarrow{AB}$ and $\overline{CD}$, construct point E on $\overrightarrow{AB}$ such that $\overline{AE} \cong \overline{CD}$.

With a ruler and modern compass, I.3 has the obvious solution in the notation above of letting E be the intersection of $\overrightarrow{AB}$ and A_{CD}. However, the trick is to construct A_{CD} using a ruler and a euclidean compass. This is what I.2 is all about and a solution is not as obvious. So I.2 is exactly what is required so that I.3 follows from one more use of a euclidean compass. Once we have a ruler and euclidean compass solution for I.2, we will be able to conclude that the ruler and euclidean compass together are equivalent to the ruler and modern compass. We now restate I.2 to introduce some notation for the given points and then give Euclid's construction.

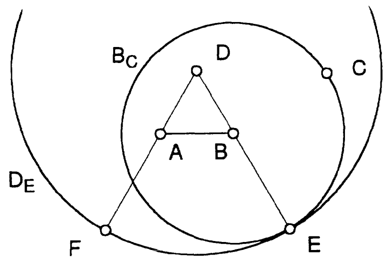
> **Problem:** Given point A and segment $\overline{BC}$, construct $\overline{AF}$ such that $\overline{AF} \cong \overline{BC}$.
> **Construction:** Given three points A,B,C, let D be a point of intersection of circles A_B and B_A. Let E be the point of intersection of B_C and $\overrightarrow{DB}$ such that B is between D and E. Let F be the point of intersection of D_E and $\overrightarrow{DA}$. Then $\overline{AF} \cong \overline{BC}$.
> **Proof:** Since $DA + AF = DF = DE = DB + BE = DA + BC$, then $AF = BC$. ■

This proof is a bare bones version. You may want to fill out the argument with more detail. Actually, the proofs of most of the constructions in this book will be left to the reader. Usually, once a construction is stated, a proof is not too difficult to find. The hard part is to find the construction in the first place, even though the problem may have several solutions. Euclid's construction for I.2 is much clearer if you look at Figure 1.1. You may have noticed a nice convention that helps us follow the details of a construction, namely that new points are usually named in alphabetical order. With this convention in mind, you should be able to write down the construction theorem by looking at the construction drawing alone.

You see the drawing and the statement of the problem in Figure 1.1. By now you are aware that the statement of a problem is not unique; you have seen I.2 in three different forms. You should be wondering what the scheme below the statement of I.2 in Figure 1.1 is all about. This is a shorthand notation for the construction stated in words above. The scheme

p, q
P, Q

is short for "Let P and Q be points of intersection of figures p and q." Associate each of the four sentences of the displayed section above that is labeled "Construction" with one of the four parts of the scheme in Figure 1.1; the last part of the scheme simply declares F to be what we were looking for. If there are any restrictions on the points, such as that B be between D and E, then these are written into the scheme as is also illustrated in Figure 1.1. For practice in understanding these schemes, Exercise 1.1 should be done now.



Euclid 1.2: Given three points A,B,C, construct point F such that $AF = BC$.

A_B, B_A	$B_C, \overrightarrow{DB}$	$D_E, \overrightarrow{DA}$	F
C	E D–B–E	F	

FIGURE 1.1.

Exercise 1.1. Draw sketches and write out in words the constructions for Euclid I.9 and I.10 given by the schemes below.

Euclid I.9. Construct the angle bisector of $\angle ABC$.

$B_A, \overrightarrow{BC}$	A_B, D_B	$\overrightarrow{BE}$
D	B, E	

Euclid I.10. Construct M, the midpoint of $\overline{AB}$.

A_B, B_A	$\overline{AB}, \overline{CD}$	M
C, D	M	

For practice in making the schemes, Exercise 1.2 should be done at this time.

Exercise 1.2. Give constructions in the form of a scheme for Euclid I.11 and I.12.

Euclid I.11. Given $\overleftrightarrow{AB}$, construct $\overleftrightarrow{AD}$ such that $\overleftrightarrow{AD} \perp \overleftrightarrow{AB}$.

Euclid I.12. Given point C off $\overleftrightarrow{AB}$, construct $\overleftrightarrow{CD}$ such that $\overleftrightarrow{CD} \perp \overleftrightarrow{AB}$.

Check your answers to Exercises 1.1 and 1.2 with those given in The Back of the Book. You are required to be able to read and understand these schemes. However, you are not required to use them for yourself, since you can always write out the constructions in sentences. Indeed, after the elementary constructions are assumed, it is certainly preferable to write "Let M be the midpoint of $\overline{AB}$" than to write down the scheme for the construction. The schemes are handy but they should be used with some discretion.

The structure of this first chapter is rather informal. For each of Euclid's problems, you may want to write down the statement of the problem in detail, to state a formal construction, to give a proof of that construction theorem, and to execute a construction drawing. On the other hand, for most problems, you may be happy with only a sketch and/or an outline of a construction. In any case, the development in this chapter should be considered to be merely an outline. Only an indication of a construction will be given for some problems. The important thing is that you add as much to the outline as is required for an understanding of the euclidean constructions and as you enjoy doing.

Euclid I.23. Given $\overline{AB}$ and $\angle CDE$, construct $\angle BAH$ congruent to $\angle CDE$.

$D_{AB}, \overrightarrow{DC}$	$D_{AB}, \overrightarrow{DE}$	A_B, B_{FG}	$\angle BAH$
F	G	H	

Euclid I.23 poses the problem of "copying" a given angle, that is, to construct an angle congruent to the given angle. The scheme above gives a theorem that solves the problem as stated. Since we can also copy segments by I.3, then we can copy triangles. It follows that with a ruler and a compass we can construct a polygon congruent to any given polygon.

Euclid I.31. Given point P off $\overleftrightarrow{AB}$, construct the line through P that is parallel to $\overleftrightarrow{AB}$.

Given P off $\overleftrightarrow{AB}$, by the construction for I.23 we can construct $\angle APQ$ such that $\angle APQ$ and $\angle PAB$ are congruent alternate interior angles for lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{PQ}$ with transversal $\overleftrightarrow{AP}$. This provides a construction for I.31 whose proof is immediate. However, this is only one construction for this very important problem. There are many more. We are not bound by the order of Euclid's propositions. You may use any theorem you can think of from plane geometry to develop a new construction for I.31 in the following exercise. Again, you are requested to do the exercise at this time. This warming-up exercise may prove to be time well spent, even if you do not produce many solutions to the problem.

Exercise 1.3. Give as many constructions for Euclid I.31 as you can think of in two sessions of one hour each.

In passing, we mention that the existence and uniqueness of the line in I.31 is often mistakenly called Euclid's parallel postulate. As an axiom, it should be called *Playfair's Parallel Postulate*, since John Playfair (1748–1819) suggested using the proposition in place of that actually used by Euclid. For the record, we state *Euclid's Parallel Postulate*: If points A and D are on the same side of $\overleftrightarrow{BC}$ and $m\angle ABC + m\angle BCD < 180$, then

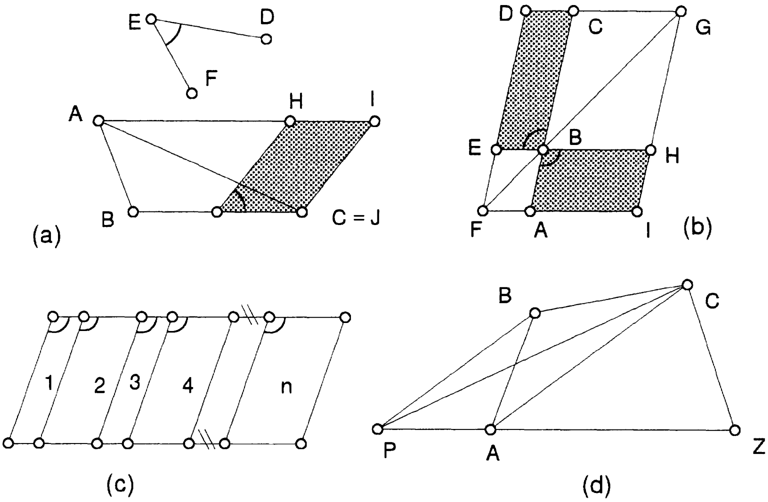


FIGURE 1.2.

$\overrightarrow{BA}$ and $\overrightarrow{CD}$ intersect. You can see why this more complicated proposition is usually replaced by some equivalent axiom.

Given $\angle ABC$, if D is the point of intersection of the line through A that is parallel to $\overleftrightarrow{BC}$ and the line through C that is parallel to $\overleftrightarrow{AB}$, then $\square ABCD$ is a parallelogram. So parallelograms with a given angle are easily constructed with two applications of a construction for I.31. Euclid I.42 requires that the constructed parallelogram also have a prescribed area, and I.44 then imposes the additional requirement that the parallelogram have a given side. Sometimes a sketch is the best way to convey the principal idea of a construction. This is the case with Euclid I.42 and Figure 1.2a, where $BG = GC$ and $\overleftrightarrow{AH} \parallel \overleftrightarrow{BC}$.

Euclid I.42. Given $\triangle ABC$ and $\angle DEF$, construct parallelogram $\square GHIJ$ such that $\angle JGH \cong \angle DEF$ and $GHIJ = ABC$.

Euclid I.44. Given $\overline{AB}$, $\triangle PQR$, and $\angle STU$, construct parallelogram $\square ABHI$ such that $\angle ABH \cong \angle STU$ and $ABHI = PQR$.

Construction for I.44. Given $\overline{AB}$, $\triangle PQR$, and $\angle STU$, let $\square BCDE$ be a parallelogram such that A–B–C, $\angle CBE \cong \angle STU$, and $BCDE = PQR$. Let F be such that $\square ABEF$ is a parallelogram. Let $\overrightarrow{FB}$ intersect $\overrightarrow{DC}$ at G. Let H be such that $\square BCGH$ is a parallelogram. Let I be such that $\square ABHI$ is a parallelogram. Then $\square ABHI$ is a parallelogram such that $\angle ABH \cong \angle STU$ and $AHIB = PQR$.

Proof. Parallelogram $\square BCDE$ is constructed by I.42 and the remaining parallelograms are constructed by I.31. See Figure 1.2b. We have $\angle ABH \cong \angle CBE \cong \angle STU$. Also, since $\overline{FG}$ is a diagonal of parallelogram $\square FDGI$, we have $ABHI = FIG - FAB - BHG = FDG - FEB - BCG = BCDE = PQR$. ■

The use of previously proved propositions to prove a new proposition characterizes mathematics. Almost every mathematical proof relies on previously proved propositions, which are not reproved when they are needed in a new argument. The proof above refers to previous constructions without repeating them. Only in the case of a drawing are we required to go back and retrace old steps. The construction and its proof may have suggested to you a problem that we will mention here and solve in the next chapter, where we are necessarily more formal. The problem is that of distinguishing from all the points in the plane those points that we know we can construct with a ruler and a compass. It is implicit in the statement of our constructions that all the points mentioned can be constructed.

Euclid I.45. Construct a parallelogram having a given side, a given angle, and an area equal to that of a given polygon.

The idea for I.45 is to triangulate the given polygon into n triangles and then construct parallelogram $\square ABZY$ by n repeated applications of I.44, as suggested by Figure 1.2c. This should be sufficient consideration for I.45. In the best tradition of mathematical economy, this construction and its proof depend on previously proved constructions. Nothing that is essentially new has to be introduced. On the other hand, if we actually have to carry out a construction for I.45, then some short cuts would be very welcome. A different approach to I.45 replaces the repeated application of I.44 by the repeated application of a simpler construction. The idea is to repeat producing a polygon having the same area as the given polygon but, until a triangle is obtained, with one less side than the previous polygon. To indicate the technique, consider the given $\square ABCZ$ in Figure 1.2d. Point P is constructed as the intersection of $\overleftrightarrow{AZ}$ and the line through B that is parallel to $\overleftrightarrow{AC}$. Then $\triangle APC$ and $\triangle ABC$ have a common base $\overline{AC}$ and the altitudes to this base are congruent. So $APC = ABC$, which implies $ZPC = ABCZ$, as desired. Once a triangle is obtained, then I.45 reduces to one application of I.44.

Euclid I.46 is the last construction problem in Book I. The first Book then ends with the Pythagorean Theorem (I.47) and its converse (I.48). Doing I.46 economically is left for Exercise 1.11.

Euclid I.46. Given $\overline{AB}$, construct square $\square ABCD$.

It is generally supposed that deductive mathematics began with Thales of Miletus about 600 BC, which is three hundred years before Euclid. The details of which propositions the first true mathematician proved first and with what arguments are lost to history. Nevertheless, we will refer to a most likely candidate for the title of "the world's oldest theorem" as the *Theorem of Thales*: An angle inscribed in a semicircle is a right angle. We shall use the Theorem of Thales often.

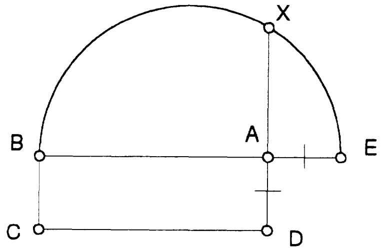


FIGURE 1.3.

Since we are not studying an axiomatic development of geometry, we are not bound by Euclid's order of the propositions. We will consider II.11 after II.14 because II.14 is the goal of the sequence we have been looking at. Euclid II.14 completes the logical sequence of propositions I.23, I.42, I.44, I.45, and I.46.

Euclid II.14. Construct a square with an area equal to that of a given polygon.

Euclid's construction for II.14 begins with the construction by I.45 of a rectangle $\square ABCD$ having area equal to the area of the given polygon. See Figure 1.3. Let E be the point such that B–A–E and $AE = AD$. Let the perpendicular to $\overline{AB}$ at A intersect the circle with diameter $\overline{BE}$ at X. Since $\angle BXE$ is right by the Theorem of Thales, then $\angle ABX \cong \angle AXE$. It follows that $\triangle AXE \cong \triangle ABX$ and, hence, that $AX/AE = AB/AX$. So $(AX)^2 = (AB)(AE) = (AB)(BC)$. Thus, a square constructed by I.46 on side $\overline{AX}$ has area $ABCD$, as desired.

Specific problems may have special solutions. After just considering II.14, which depends on the complicated I.45, you might be tempted to give the "wrong" solution to the following problem: Square two given squares. To "square a polygon" requires the construction of a square having the area of the given polygon. Euclid II.14 completed a sequence of propositions showing how to square any given polygon with ruler and compass. The concept is easily generalized to squaring any given region or regions. To square two given squares, you are required to construct one square having the area equal to that of the sum of the areas of the two given squares. For this particular problem, you should be able to give a special solution that is far better than one suggested by I.45 and II.14. In the next chapter, we will discuss the fact that there is no ruler and compass solution to the problem of squaring a circle.

We now return to II.11, the seemingly peculiar problem of "cutting $\overline{AB}$ in extreme and mean ratio." This means we are to find the point X on $\overline{AB}$ such that $AB/AX = AX/BX$. Then $(AB)(BX) = AX^2$. (There should be no confusion when for points P and Q we write PQ^2 in place of the more formal $(PQ)^2$, particularly since "Q^2" does not make any sense.) Although we have no motivation for the problem at this time, we shall soon find out

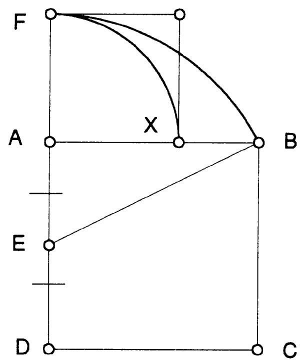


FIGURE 1.4.

that solving this problem is just what is needed in order to construct a regular pentagon.

Euclid II.11. Given $\overline{AB}$, construct point X on $\overline{AB}$ such that $(AB)(BX) = AX^2$.

Construction for II.11. Given $\overline{AB}$, let C and D be such that $\square ABCD$ is a square. Let E be the midpoint of $\overline{AD}$. Let E_B intersect $\overrightarrow{EA}$ at F. Let A_F intersect $\overrightarrow{AB}$ at X. Then X is on $\overrightarrow{AB}$ and $(AB)(BX) = AX^2$.

Proof. Points C and D are constructed by I.46. See Figure 1.4. Point E is constructed by I.10. Points F and X are constructed as the intersection of constructible circles and constructible lines. Since $AX+AE = EF = EB < AB + AE$, then $AX < AB$ and so X is on $\overline{AB}$. Since $AB^2 + AE^2 = BE^2$, then $AB^2 = BE^2 - AE^2 = EF^2 - AE^2$. So

$$\begin{aligned}(AB)(BX + AX) &= AB^2 = EF^2 - AE^2\\ &= (EF + AE)(EF - AE) = (FD)(AF)\\ &= (AX + AB)(AX).\end{aligned}$$

Then, multiplying out the right-hand side and the left-hand side of the string of equalities, we have $(AB)(BX)+(AB)(AX) = (AX)^2+(AB)(AX)$. Hence, $(AB)(BX) = AX^2$. ∎

Now that we have cut $\overline{AB}$ in extreme and mean ratio, let's see what arithmetic value these ratios AB/AX and AX/BX have. Let $AB = a$ and $AX = x$. Then we have $a(a - x) = x^2$ or, after we rearrange and divide both sides by a^2, the quadratic equation $(x/a)^2 + (x/a) - 1 = 0$. Solving the quadratic equation by using the quadratic formula, we have

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

In Figure 1.4, we have $AF/AD = AX/AB = x/a = (-1 + \sqrt{5}\,)/2$, which will be useful later. Perhaps you recognize the ratio a/x as the *golden section*, which is related to the Fibonacci sequence. That a segment of length

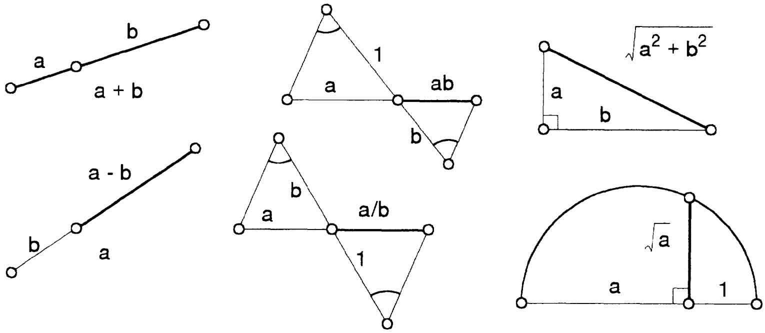


FIGURE 1.5.

the golden section or its reciprocal can be constructed also follows from the segment arithmetic that we interpose here in the form of Figure 1.5. Specifically, given a unit segment and segments of lengths a and b, we can construct segments of length $a + b$, $a - b$ when $a > b$, ab, a/b, $\sqrt{a^2 + b^2}$, and $\sqrt{a}$. Note that the square root construction is contained in Figure 1.3 with $AB = a$ and $AE = 1$

Euclid III.17. Construct a tangent to a given circle through a given point outside the circle.

Construction for III.17. Suppose O–A–P. Let the perpendicular to $\overleftrightarrow{OP}$ at A intersect O_P at B. Let $\overline{OB}$ intersect O_A at Q. Then $\overleftrightarrow{PQ}$ is tangent to O_A at Q.

Although the construction above is from the *Elements*, it has a modern flavor. That $\overleftrightarrow{PQ}$ is a tangent to O_A at Q follows because $\overleftrightarrow{PQ}$ is the image of constructed tangent $\overleftrightarrow{AB}$ to O_A at A under the reflection in the angle bisector of $\angle AOB$. However, Euclid's construction for a tangent may not be the one you are most familiar with. A different, better known construction uses Euclid III.31, which is the Theorem of Thales. See Figure 1.6 for a sketch of the two constructions. The details of the proofs are your responsibility.

Euclid III.36 and III.37 are theorems that you may have forgotten. They are stated below; they are not constructions. Euclid III.37, which we will need when we get to IV.10, is both a converse and a consequence of III.36. In proving III.36, use two facts: (1) the measure of the angle between a chord of a circle and the tangent at one end of the chord is half that of the intercepted arc; and (2) the measure of an angle inscribed in a circle is half that of its intercepted arc. So $\triangle PQR \sim \triangle PSQ$ in the following statement of III.36. The remaining part of the proofs are left for you.

Euclid III.36. If Q,R,S are points on a circle with tangent $\overleftrightarrow{PQ}$ and if secant $\overleftrightarrow{RS}$ passes through P, then $(PR)(PS) = PQ^2$.

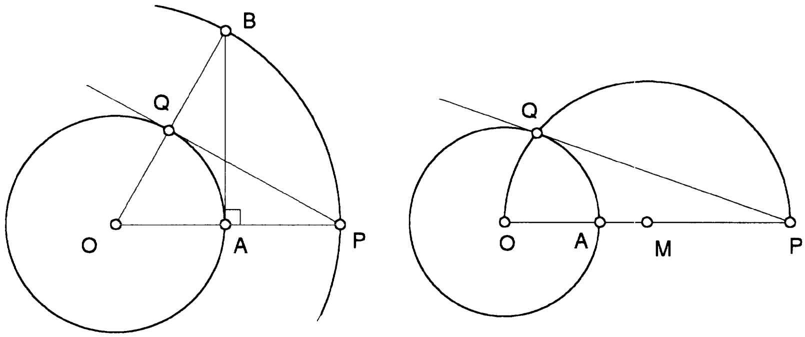


FIGURE 1.6.

Euclid III.37. If Q,R,S are points on a circle, point P is outside the circle, secant $\overleftrightarrow{RS}$ passes through P, and $(PR)(PS) = PQ^2$, then $\overleftrightarrow{PQ}$ is a tangent to the circle.

We turn to the constructions from Book IV of the *Elements* after mentioning only the following two easy constructions from later Books. Figure 1.7a illustrates VI.12. Here, we have $p/q = r/x$ and so $x = rq/p$. Hence, by taking p or q to be 1, we see that VI.12 essentially gives us the product of given lengths r and q as well as the quotient of given lengths r and p. Figure 1.7b illustrates VI.13, which echoes II.14. Here, we have $p/x = x/q$ and so $x = \sqrt{pq}$. Hence, we see that VI.13 gives us $\sqrt{pq}$, the geometric mean of given lengths p and q.

Euclid VI.12. Construct a fourth proportional to three given segments. (Given segments of lengths p,q,r, then construct a segment of length x such that $p/q = r/x$.)

Euclid VI.13. Construct a mean proportional to two given segments. (Given segments of lengths p and q, then construct a segment of length x such that $p/x = x/q$.)

We are going to jump over the first nine constructions from Book IV. Read the statement below of IV.10 and try to think why you might want to solve this problem.

Euclid IV.10. Construct an isosceles triangle having base angles that are double the third angle.

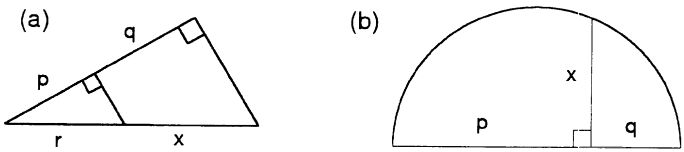


FIGURE 1.7.

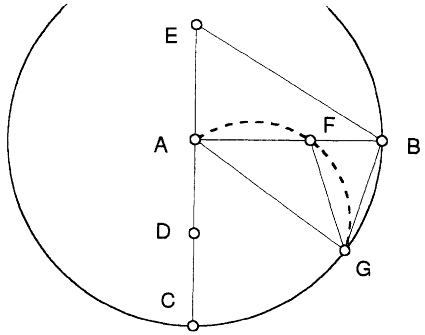


FIGURE 1.8.

Construction for IV.10. Given A_B, let the perpendicular to $\overline{AB}$ at A intersect A_B at C. Let D be the midpoint of $\overline{AC}$. Let D_B intersect $\overrightarrow{CA}$ at E, and let A_E intersect $\overline{AB}$ at F. Let B_{AF} intersect A_B at G. Then $m\angle BAG = 36$, and $m\angle ABG = m\angle AGB = 2\,m\angle BAG = 72$.

Proof. By Euclid II.11, point F is constructed on $\overline{AB}$ such that $(AB)(BF) = AF^2 = BG^2$. See Figure 1.8. Point G is constructed such that $BG = AF$. Since $BG^2 = (BF)(BA)$, then $\overleftrightarrow{BG}$ is tangent to the circumcircle of $\triangle AFG$. (Here is where we use III.37, which is given above.) Hence, $\angle BAG \cong \angle FGB$. Then, by the Exterior Angle Theorem applied to $\triangle GAF$, we have $m\angle BFG = m\angle BAG + m\angle FGA$. So $\angle BFG \cong \angle BGA$. Since $\triangle ABG$ is isosceles, then $\angle FBG \cong \angle BGA$. Hence, $\triangle GBF$ is isosceles and $GF = BG = AF$. So, $\angle FGA \cong \angle FAG$. Then $m\angle BGA = 2\,m\angle BAG$. Therefore, $m\angle BAG = 36$, as desired. ■

We now know how to construct a regular decagon, since we can construct an angle of 36°. In Figure 1.8, we can march along the circumference of A_B marking off vertices of distance BG each from another. Taking alternate vertices of the regular decagon, we have the vertices of a regular pentagon. This would be a great construction for a regular pentagon if there were not an even better one that we have already done without realizing it. We will show that $\overline{BE}$ in Figure 1.8 is congruent to a side of a regular pentagon inscribed in A_B. Because $BG = AF = AE$, we already know $\overline{AE}$ is congruent to a side of a regular decagon inscribed in A_B.

Definition. Let s_n denote the length of a side of a regular n-gon inscribed in a circle having unit radius.

Consider a regular n-gon inscribed in a circle with radius 1. With the perpendicular drawn from the center to one side, it follows from elementary trigonometry that $\sin(360/2n)^\circ = (s_n/2)/1$. Hence, $s_n = 2\sin(180/n)^\circ$. We have already determined that $s_{10} = AE/AB = (-1+\sqrt{5}\,)/2$ in Figure 1.8. Since $s_{10} = 2\sin 18^\circ = 2\cos 72^\circ$, then $\cos 72^\circ = s_{10}/2 = (-1+\sqrt{5}\,)/4$. By the half-angle formula $\sin(x/2) = \sqrt{(1-\cos x)/2}$ from trigonometry, we

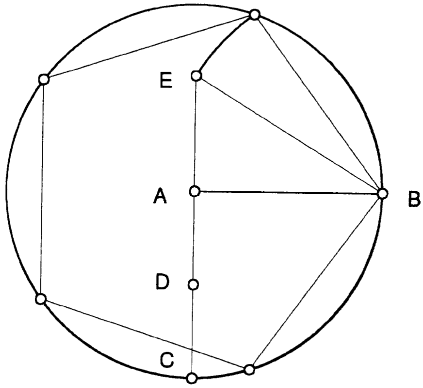


FIGURE 1.9.

have $\sin 36° = \sqrt{(1 - \cos 72°)/2}$. Therefore,

$$s_5 = 2 \sin 36° = 2\sqrt{(1 - \cos 72°)/2} = \sqrt{(5 - \sqrt{5}\,)/2}.$$

We put this together with some results that are easier to get:

Theorem. In general, $s_n = 2 \sin(180/n)°$. In particular,

$$s_3 = \sqrt{3}, \quad s_4 = \sqrt{2}, \quad s_5 = \sqrt{\frac{5 - \sqrt{5}}{2}}, \quad s_6 = 1, \quad s_{10} = \frac{-1 + \sqrt{5}}{2}.$$

You may be less than excited about finding s_5. However, the remarkable thing about our calculation of s_5 is that it is now just a trivial calculation in arithmetic to prove the following theorem of Euclid.

Euclid XIII.10.

$$s_5^2 = s_6^2 + s_{10}^2.$$

In Figure 1.8, we have $s_6 = AB/AB$, $s_{10} = AE/AB$, and $\triangle ABE$ is a right triangle. Hence, from XIII.10 and the Pythagorean Theorem, we see that $s_5 = BE/AB$ in the figure. The points F and G are used for the proof of the construction for IV.10 but are not needed for the construction of $\overline{BE}$. Therefore, as a consequence of XIII.10, we have the promised shorter construction that leads to a regular pentagon. This is stated below. See Figure 1.9, where we have $BE/AB = s_5$ and $AE/AB = s_{10}$. That is, $\overline{BE}$ is congruent to a side of a regular pentagon inscribed in A_B, and $\overline{AE}$ is congruent to a side of a regular decagon inscribed in A_B.

Construction for a regular pentagon and a regular decagon (Euclid). Let the perpendicular to $\overline{AB}$ at A intersect A_B at C. Let D be the midpoint of $\overline{AC}$. Let D_B intersect $\overrightarrow{CA}$ at E. Then $\overline{BE}$ is congruent to a side of a regular pentagon inscribed in A_B, and $\overline{AE}$ is congruent to a side of a regular decagon inscribed in A_B .

A pentadecagon is a 15-gon. We offer the following construction for a regular triangle (an equilateral triangle) inscribed in a given circle O_P. Check it out.

O_P, P_O	A_B, O_P	$\triangle ABC$
A, B	B, C	

Since we can construct an equilateral triangle and a regular pentagon with ruler and compass, it follows that we can construct angles of 60° and 36°. The existence of a ruler and compass construction for a regular pentadecagon now follows from a simple theorem of arithmetic: $360/15 = 24 = 60 - 36$.

Euclid IV.16. In a given circle, inscribe a regular pentadecagon.

If a regular n-gon can be constructed with a ruler and a compass, then a regular $2n$-gon can be constructed with a ruler and a compass. We can now construct all the regular polygons that the ancient Greeks could construct with a ruler and a compass:

The regular n-gons constructible with a ruler and a compass in antiquity. Values for n:

$$\begin{aligned}
&3, 6, 12, 24, 48, 96, \ldots,\\
&4, 8, 16, 32, 64, 128, \ldots,\\
&5, 10, 20, 40, 80, 160, \ldots,\\
&15, 30, 60, 120, 240, 480, \ldots.
\end{aligned}$$

This is where things stood until March 29, 1796, when at age eighteen Carl Friedrich Gauss (1777–1855) created a ruler and compass construction for a regular 17-gon. Gauss's birthday was April 30 and so he was nineteen when he announced the result on the first day of June. This construction of the regular heptadecagon by the greatest mathematician since Newton (1642–1727) was published in 1801. Which regular n-gons are constructible with a ruler and a compass will be explained in the next chapter.

Herbert W. Richmond of King's College, Cambridge, gave an efficient ruler and compass construction for a regular heptadecagon in 1893 and again with more detail in 1909. We give, without proof, Richmond's construction for a regular 17-gon that is inscribed in a circle and has vertices V, V_1, V_2, V_3, ... , V_{16}. It is sufficient to find adjacent vertices V_3 and V_4 in order to determine a side of the inscribed heptadecagon. See Figure 1.10.

Construction for a regular heptadecagon (H. W. Richmond). Let O_V be a unit circle with diameter $\overline{VA}$. Let $\angle BOV$ be right with $BO =$

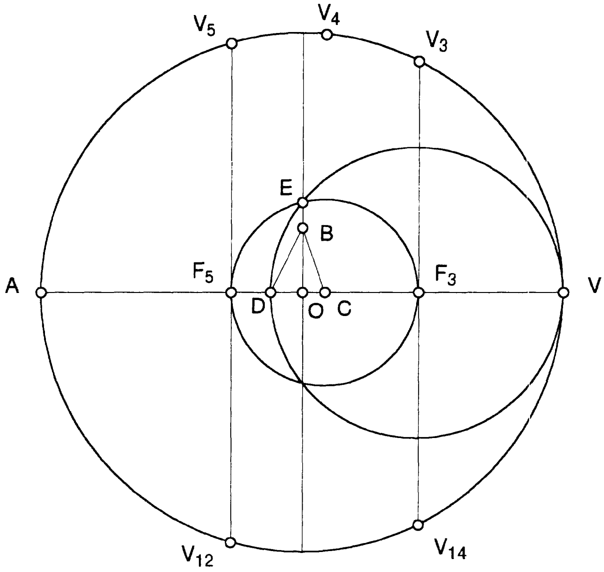


FIGURE 1.10.

$OV/4$. Let C be on $\overline{OV}$ with $m\angle OBC = (m\angle OBV)/4$. Let D be on $\overline{OA}$ with $m\angle CBD = 45$. Let the circle with diameter $\overline{DV}$ intersect $\overrightarrow{OB}$ at E. Let $\mathcal{C}_E$ intersect $\overrightarrow{CV}$ at F_3 and intersect $\overrightarrow{CA}$ at F_5. Let the perpendiculars to $\overleftrightarrow{VA}$ at F_3 and F_5 intersect a semicircle on $\overline{VA}$ at V_3 and V_5. Let V_4 be the intersection of $\mathcal{O}_V$ and the perpendicular bisector of the segment with endpoints V_3 and V_5. Then $V_3V_4 = s_{17}$.

You should do both of the following exercises now. If you can resist, do not read Exercise 1.5 until you have completed Exercise 1.4.

Exercise 1.4. Read over Richmond's construction for a regular heptadecagon until you feel familiar with it. With ruler, compass, and watch at hand, give yourself a large circle that takes up one full page. Note the time. Carry out the Richmond construction with the ruler and compass to the point of determining the seventeen vertices. Note the time when you are finished and calculate the time it took you to execute the construction. Do the construction only once, even if it "doesn't work."

Exercise 1.5. With ruler, compass, and watch at hand, give yourself a large circle that takes up one full page. Note the time. By trial and error, determine to a reasonable degree of accuracy the vertices of a regular heptadecagon inscribed in your circle. Note the time it takes to get your seventeen points. If you have not already done so, do Exercise 1.4; compare your two times and your two figures. What conclusions can you draw from these two exercises?

The rule of the game of geometric constructions that limits the tools to the ruler and the compass is called the "Platonic restriction" and predates Euclid. Although we cannot be sure what prompted Plato, this limitation does provide an elegant game. The ruler and compass constructions are constructions by means of the most basic geometric elements—the lines and the circles. However, the ruler and compass are not the only possible construction tools. We shall look at several possibilities in this book.

The one nonplatonic tool that we consider now is the *tomahawk*. The inventor of this charming construction tool is unknown. Descriptions appeared thirty years before the publication of the extensive study by Pierre-Joseph Glotin, a retired French naval officer and member of the Légion d'Honneuer from Bordeaux. Glotin's 1863 paper begins: A very ingenious pratical solution to the problem of trisecting angles employs the use of the well-known instrument named *trisector*. Glotin's ("depuis longtemps connu") trisector, which could not have been too well known, was reported in detail to the larger mathematical community in 1877 by Henri Brocard. Pierre Glotin's trisector is our tomahawk.

The essential skeleton of the tomahawk (or shoemaker's knife) is shown in Figure 1.11a, where $AB = BC = CD$, the semicircle has diameter $\overline{BD}$, and $\overrightarrow{BE} \perp \overline{BA}$. A decorative, fleshed-out version is shown in Figure 1.11b, where the use of the tomahawk is also illustrated. The tomahawk's special function is to trisect angles. The tomahawk is inserted in a given angle $\angle UVW$ so that V lies on $\overrightarrow{BE}$, point A lies on one side of the angle, and the semicircle is tangent to the other side of the angle. Now, you are to convince yourself that $\overrightarrow{VB}$ and $\overrightarrow{VC}$ do trisect the given angle.

Attaching to a tomahawk the proper rigid T, hinged at the point of the tomahawk, as indicated in Figure 1.11c, we produce an angle quinquesector that is actually a quinsector. (To quinquesect a given angle is to construct an angle having measure one-fifth the measure of the given angle; to quinsect a given angle is to divide the angle into five congruent parts. "Quinquesector" seems to have a more interesting ring to it than "quinsector" and the frequently used word "quintisector.") Attaching more T's, we would produce a construction tool that divides an angle into any odd number of congruent parts, as Glotin pointed out in 1863. An alternative method for doing this is to use several congruent tomahawks at once. For example, by attaching two congruent tomahawks together (such that at its point A one pivots about the point C of the other) as in Figure 1.11d, we can achieve the quinsection of a given angle.

Until recently, all geometers would know many of Euclid's propositions by number. Although this is no longer the case, we shall assume hereafter that the statement of each of Euclid I.2, Euclid I.3, and Euclid I.31 is known by number and we shall refer to them in this way. Euclid I.2 is important to us because the problem requires showing a modern compass and a euclidean compass are equivalent. Note that we have shown above only that they are equivalent in the presence of a ruler. We will want to find out whether the

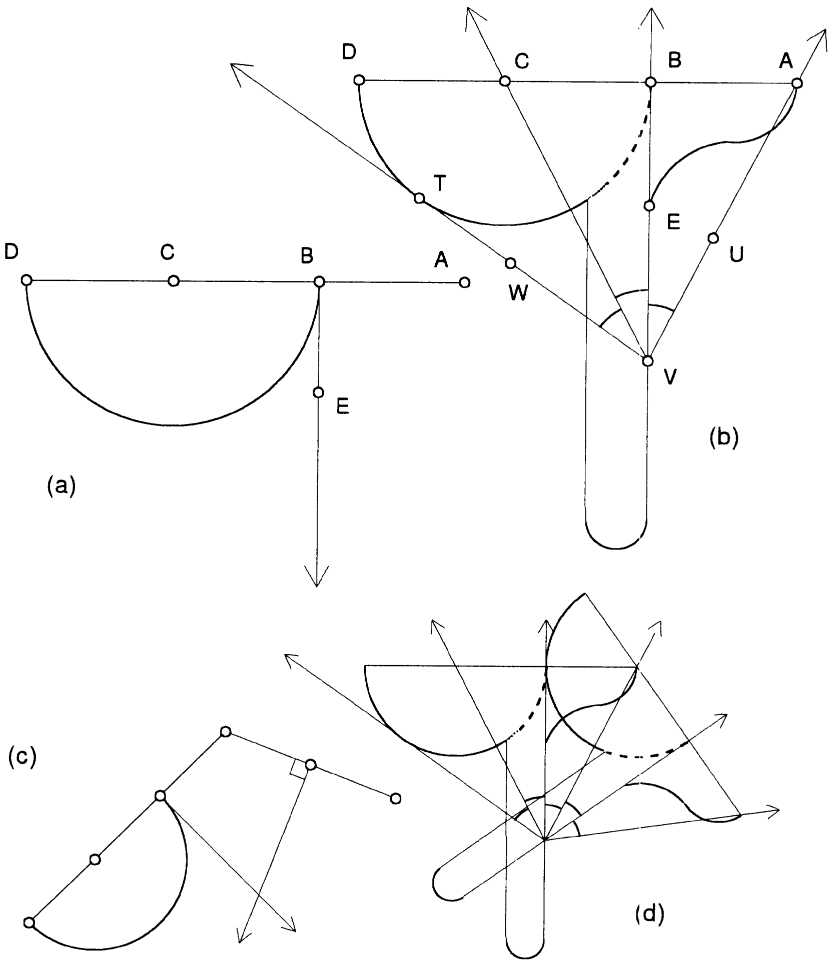
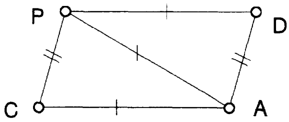


FIGURE 1.11.

ruler is necessary for this equivalence. As previously mentioned, Euclid I.3 describes what a dividers does. Later, we shall investigate what we can do with a ruler and a dividers, without a compass. Finally, Euclid I.31 is intrinsically important simply because of the importance of parallels in euclidean geometry. This proposition will turn up repeatedly and we shall want to refer to some of the particular solutions to this problem. The following catalog of constructions is by no means exhaustive. SAS, SSS, and SAA denote the familiar congruence theorems for triangles. You will be able to deduce what we are calling the Side-splitting Theorem from the fourth construction. We are constructing the line through P that is parallel to $\overleftrightarrow{AB}$, where P is off $\overleftrightarrow{AB}$.

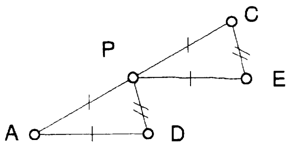
(1) Using alternate interior angles (Euclid):

$A_P, \overleftrightarrow{AB}$	P_A, A_{PC}	$\overleftrightarrow{PD}$
C	D $\overline{DC}$ intersects $\overleftrightarrow{AP}$	



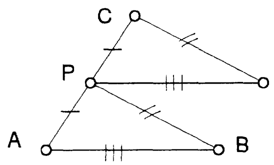
(2) Using corresponding angles:

$P_A, \overleftrightarrow{AP}$	$A_P, \overleftrightarrow{AB}$	P_C, C_{PD}	$\overleftrightarrow{PE}$
A, C	D	E $\overline{DE}$ doesn't intersect $\overleftrightarrow{AP}$	



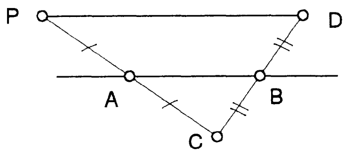
(3) Using corresponding angles and SSS:

$P_A, \overleftrightarrow{AP}$	C_{PB}, P_{AB}	$\overleftrightarrow{PD}$
A, C	D $\overline{BD}$ doesn't intersect $\overleftrightarrow{AP}$	



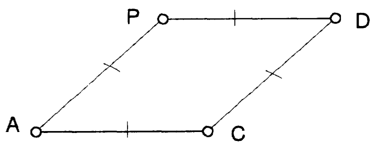
(4) Using the Side-splitting Theorem:

$A_P, \overleftrightarrow{AP}$	$B_C, \overleftrightarrow{BC}$	$\overleftrightarrow{PD}$
P, C	C, D	



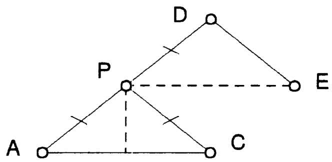
(5) Using a rhombus:

$A_P, \overleftrightarrow{AB}$	C_A, P_A	$\overleftrightarrow{PD}$
C	A, D	



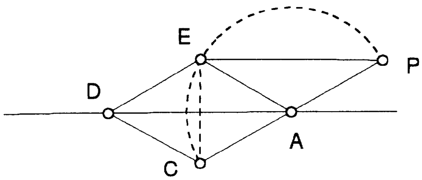
(6) $\perp$ to $\perp$, using angle bisectors of a linear pair of angles are $\perp$; take $PA \geq PB$ (so that $C \neq A$ below):

$P_A, \overleftrightarrow{AB}$	$P_A, \overleftrightarrow{AP}$	C_P, D_P	$\overleftrightarrow{PE}$
A, C	A, D	P, E	



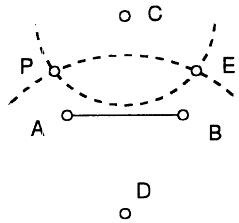
(7) $\perp$ to $\perp$, using Theorem of Thales; take $PA \geq PB$:

$A_P, \overleftrightarrow{AP}$	$C_A, \overleftrightarrow{AB}$	D_C, A_C	$\overleftrightarrow{PE}$
P, C	A, D	C, E	



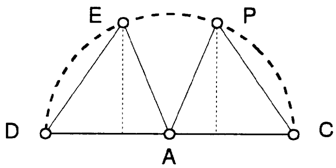
(8) Using $\perp$ to $\perp$ again; take $\perp$ bisector of $\overline{AB}$ off P:

A_B, B_A	C_P, D_P	$\overleftrightarrow{PE}$
C, D	P, E	



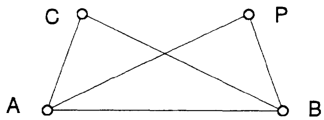
(9) Equidistant curve; take $PA \geq PB$:

$A_P, \overleftrightarrow{AB}$	A_P, D_{CP}	$\overleftrightarrow{PE}$
C, D	E $\overline{PE}$ doesn't intersect $\overleftrightarrow{AB}$	



(10) Short but requires having $\perp$ bisector of $\overline{AB}$ off P:

A_{BP}, B_{AP}	$\overleftrightarrow{PC}$
C $\overline{PC}$ doesn't intersect $\overleftrightarrow{AB}$	



(11) Quick and dirty:

P_{AB}, B_{AP}	$\overleftrightarrow{PC}$
C $\overline{AC}$ intersects $\overleftrightarrow{BP}$	

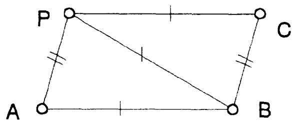


FIGURE 1.12.

One of the amazing successes of secondary education is the respect that is learned for mathematics. It is not unusual for a student who has done Exercises 1.4 and 1.5 to compare a pathetic curve that doesn't even form a polygon but was drawn with a ruler and compass to an apparently perfect regular 17-gon that was drawn by trial and error and then to conclude that the ruler and compass construction is "the better 17-gon because it is—or could be—more accurate," in spite of the presence of the obvious contradictory evidence. On the other hand, this faith does have its foundation. The pathetic curve does represent an interesting theorem, while the handsome polygon is just that, a drawing and nothing more. If we want a drawing of a regular 17-gon, we do not go to Richmond's construction for directions. Exaggeration often helps to see things more clearly. Would we seriously consider following the very lengthy directions, which do exist(!), for a ruler and compass construction of a regular 257-gon? We know beforehand that our result would most likely be way off. Our experience has taught us that the accuracy of even a very simple construction drawing is a function of how recently we sharpened our pencil. The point of all this is that we are interested in the mathematical theory suggested by the tools. If not, then we should do something else instead of studying geometric constructions. Geometric constructions is an interesting mental game. Our imperfect construction drawings can only suggest our perfect construction theorems.

Exercises

1.1. Draw sketches and write out in words the constructions for Euclid I.9 and I.10 given by the schemes below.◊

Euclid I.9: Construct the angle bisector of $\angle ABC$.

$B_A, \overrightarrow{BC}$	A_B, D_B	$\overrightarrow{BE}$
D	B, E	

Euclid I.10: Construct M, the midpoint of $\overline{AB}$.

A_B, B_A	$\overline{AB}, \overline{CD}$	M
C, D	M	

1.2. Give constructions in the form of a scheme for Euclid I.11 and I.12.◊
Euclid I.11: Given $\overleftrightarrow{AB}$, construct $\overleftrightarrow{AD}$ such that $\overleftrightarrow{AD} \perp \overleftrightarrow{AB}$.
Euclid I.12: Given point C off $\overleftrightarrow{AB}$, construct $\overleftrightarrow{CD}$ such that $\overleftrightarrow{CD} \perp \overleftrightarrow{AB}$.

1.3. Give as many constructions for Euclid I.31 as you can think of in two sessions of one hour each.

1.4. Read over Richmond's construction for a regular heptadecagon until you feel familiar with it. With ruler, compass, and watch at hand, give yourself a large circle that takes up one full page. Note the time. Carry out the Richmond construction with the ruler and compass to the point of determining the seventeen vertices. Note the time when you are finished and calculate the time it took you to execute the construction. Do the construction only once, even if it "doesn't work."

1.5. With ruler, compass, and watch at hand, give yourself a large circle that takes up one full page. Note the time. By trial and error, determine to a reasonable degree of accuracy the vertices of a regular heptadecagon inscribed in your circle. Note the time it takes to get your seventeen points. If you have not already done so, do Exercise 1.4; compare your two times and your two figures. What conclusions can you draw from these two exercises?

1.6. Give sketches for Euclid IV.2 and IV.3.◊

1.7. Give constructions and sketches for Euclid IV.4 and IV.5.◊

1.8. Give constructions and construction drawings for Euclid IV.6 and IV.15.

1.9. In a page size circle, construct an inscribed regular pentagon and an inscribed regular pentadecagon.

1.10. Given $\overline{AB}$, construct E and F such that $\square ABEF$ is a square.◊

1.11. Square the pentagon in Figure 1.13. That is, you are to draw a larger pentagon that is similar to the pentagon shown and then construct a square having the same area as your pentagon.◊

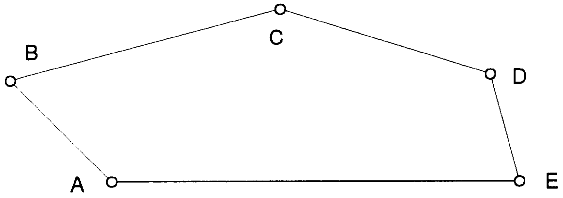


FIGURE 1.13.

1.12. Make a tomahawk. Trisect an angle with your tomahawk and prove that the construction is correct. (As you cannot draw circles having very small radii with a real compass, so a real tomahawk also has physical limitations.)

1.13. What is the area of a square inscribed in a unit circle? What is the area of a regular dodecagon inscribed in a unit circle? What is the area of a regular n-gon inscribed in a unit circle?◊

1.14. Verify Hippocrates's squaring of a lune by showing that the square with diameter $\overline{AC}$ in Figure 1.14 has the same area as the lune bounded by the arcs of M_A and C_A that are shown in the figure.

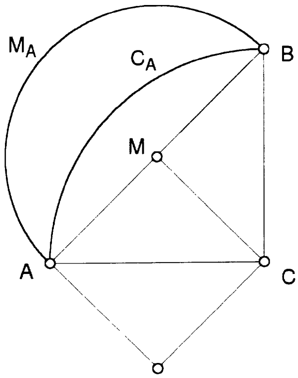


FIGURE 1.14.

1.15. Explain the use in Figure 1.15 of a carpenter's square to determine the length BP of a side of an inscribed regular pentagon in the given circle.◊

1.16. Verify the following geometric solution of the quadratic equation $x^2 - gx + h = 0$ with $h \neq 1$ if $g = 0$. The real roots are given by the intersection of the X-axis and the circle with diameter having endpoints $(0,1)$ and (g,h).◊

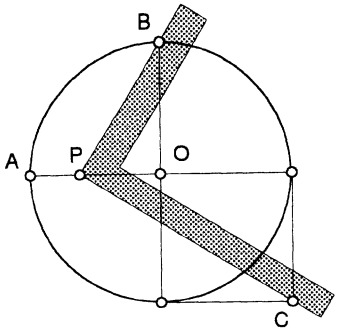


FIGURE 1.15.

1.17. Leonardo da Vinci was fascinated by squaring curved regions. Square a pair of Leonardo's claws, shown in Figure 1.16, where a 90° arc of a circle is folded over and a smaller circle is symmetrically inscribed.◊

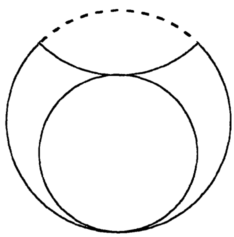


FIGURE 1.16.

1.18. Verify the first step in Archimedes' squaring of a parabola, as indicated in Figure 1.17, where $\overleftrightarrow{AS}$ and $\overleftrightarrow{BS}$ are tangents to the parabola with equation $x^2 = ky$ at points A and B on the parabola. The claim is that the area of the parabolic section cut by $\overline{AB}$ is two-thirds of the area of $\triangle ABS$. Start by using calculus to show that $S = ((a+b)/2, ab/k)$ if $A = (a, a^2/k)$ and $B = (b, b^2/k)$.

1.19. Outline the steps you would perform to square two given regular hexagons.◊

1.20. Construct a square each of whose extended sides passes through one of four given points.◊

1.21. The Problem of Apollonius (A Term Project): Construct the circles that are tangent to three given circles.◊

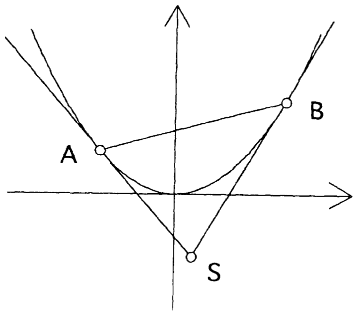


FIGURE 1.17.

$$16\cos(360/17)^\circ = -1+\sqrt{17}+\sqrt{34-2\sqrt{17}} \\ +2\sqrt{17+3\sqrt{17}-\sqrt{34-2\sqrt{17}}-2\sqrt{34+2\sqrt{17}}}$$

2
The Ruler and Compass

> The methods of coordinate geometry allow us to translate any geometric statement into the language of algebra, and though this language is less elegant, it has a larger vocabulary.
>
> HILDA HUDSON

Through his oracle at Delos, Apollo informed the Delians that if they wanted to be rid of the plague they must construct a new cubical altar that exactly doubled the volume of the existing one. *The Delian Problem* then was to construct a cube having a side $\sqrt[3]{2}$ times as long as a side of the original cube. This problem also has the somewhat misleading name *The Duplication of the Cube*. According to another legend, Eratosthenes reported that the problem was sent to the geometers at Plato's Academy in Athens. Plato is reported to have said that the god had assigned the task to shame the Greeks for their neglect of mathematics and their contempt for geometry. It was not that the Greeks could not construct segments of the required length by various methods but that they could not do so using only the ruler and compass. That was their task. There is little doubt that the Greeks soon suspected the problem had no solution. However, they lacked the algebra to prove this fact. Our task is to prove the ancient Greeks necessarily failed because they were asking for the impossible. To do this, we must formulate our problems in the language of algebra.

If we want to know what we can construct with the ruler and compass, we have to make precise what we mean by "construct." We must express in mathematical language just what it is we can do with our construction tools. With the ruler, we can draw a line provided we have two points of the line given. A ruler is sometimes called a straightedge in order to emphasize that a ruler can be used only to draw the line through two given points and cannot be used to measure length. This is a rule of the game. A ruler with marks on it is quite another game, and we shall consider that game later. With a euclidean compass, we can draw a circle provided we are given the center and a point on the circle. It is much easier to state the formal definitions in terms of a euclidean compass, rather than the modern compass that we use. From Euclid I.2, we know that, in the presence of the ruler, this euclidean compass will be equivalent to the modern compass, allowing us to draw a circle if we are given the center and any segment having the length of a radius. We have an advantage unavailable to the Greeks, namely that of coordinate geometry. The setting for all of our investigations will be the familiar x-y-plane, which is usually called the *cartesian plane* after the great French mathematician and philosopher René Descartes (1596–1650), who developed the foundation of analytic geometry.

We want to leave nothing to chance. Therefore, in order to keep track of what is going on, we do not allow ourselves to select an "arbitrary point" at any time; any point selected must have already been constructed. On the other hand, we must have two given points in order to use either the ruler or the euclidean compass in the first place. It follows that we must necessarily have a *starter set* consisting of at least two points from which we can then construct new points with our tools. We want to give a description of all these additional points that we could theoretically construct. New points are obtained as the intersections of lines and circles that we can construct from the points we have at any given time. Such a point must be a point of: (i) the intersection of two constructed lines; (ii) the intersection of a constructed line and a constructed circle; or (iii) the intersection of two constructed circles. Read the following definition of a ruler and compass constructible point carefully. Determine that the definition does model what it is that we can do with the ruler and euclidean compass to form new points from the starter set consisting of the two points $(0,0)$ and $(1,0)$.

Definition 2.1. In the cartesian plane, a point is a ***ruler and compass point*** if the point is the last of a finite sequence $P_1, P_2, \ldots, P_n$ of points such that each point is in $\{(0,0),\ (1,0)\}$ or is obtained in one of three ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; (ii) as a point of intersection of a line through two points that appear earlier in the sequence and of a circle through an earlier point and having an earlier point as center; and (iii) as a point of intersection of two circles, each of which passes through an earlier point in the sequence and each of which has an earlier point as center. A ***ruler and compass line*** is a line that passes through two ruler and compass points. A ***ruler and compass circle*** is a circle through a ruler and compass point with a ruler and compass point as center. A number x is a ***ruler and compass number*** if $(x, 0)$ is a ruler and compass point.

We point out a convention of which some beginning mathematics students are unaware. Within a definition, the word "if" usually has the meaning of "iff" that is, "if and only if." For example, in the last sentence of Definition 2.1 the statement "$(x, 0)$ is a ruler and compass point if x is a ruler and compass number" is implicitly assumed, as well as the explicit statement "if $(x, 0)$ is a ruler and compass point, then x is a ruler and compass number." It must be strongly emphasized that this convention holds only within a definition, however. From a general statement such as "$(x, 0)$ is a cartesian point if $(x, 0)$ is a ruler and compass point," it would be unwise to assume that "$(x, 0)$ is a ruler and compass point if $(x, 0)$ is a cartesian point" is true. We also mention that caution is advised in interpreting "is" as "equals." We do not make this mistake in everyday language—we know "a cat is an animal" and "an animal is a cat" are different statements—and we should not make this mistake in our mathematics.

The following six sequences are examples that satisfy the condition on $P_1, P_2, P_3, \ldots, P_n$ in the definition of a ruler and compass point above. (It is usually very difficult to check that a sequence of points does satisfy the condition.)

$$\begin{array}{l}
(1,0);\\
(1,0),(0,0),(2,0);\\
(0,0),(1,0),(-1,0),(0,\sqrt{3});\\
(0,0),(1,0),(2,0),(-2,0),(0,\sqrt{5});\\
(0,0),(1,0),(1/2,\sqrt{3}/2),(1/2,-\sqrt{3}/2),(1/2,0);\\
(0,0),(1,0),(2,0),(1/4,\sqrt{15}/4),(1/2,0).
\end{array}$$

Further, we can concatenate two or more such sequences to form a new sequence that also satisfies the condition. For example, form a new sequence by tacking the last sequence listed above to the end of the fourth sequence, as follows:

$$(0,0),(1,0),(2,0),(-2,0),(0,\sqrt{5}),(0,0),(1,0),(2,0),(1/4,\sqrt{15}/4),(1/2,0).$$

It does not matter that some points appear more than once in this concatenate sequence, which is easily seen to satisfy the required condition that each point is in $\{(0,0),\ (1,0)\}$ or is obtained in one of the three ways denoted as (i), (ii), and (iii) in Definition 2.1.

Generally, a ruler and compass point, as defined above, would be called simply a constructible point. Since we are going to be constructing points by means of other construction tools later on, this would lead to confusion. It is true however that if nothing is said to indicate the contrary, "constructible" usually means by the classical tools of ruler and compass. We confess to using informal abbreviations such as "r-c-point" in place of the formal "ruler and compass point" in conversation; you can decide whether or not you should allow yourself to write such things as, "That line is rc."

We next check that the definition above does give the desired fundamental intersection properties of ruler and compass lines and ruler and compass circles.

Theorem 2.2. *The point of intersection of two ruler and compass lines is a ruler and compass point; a point of intersection of a ruler and compass line and a ruler and compass circle is a ruler and compass point; and a point of intersection of two ruler and compass circles is a ruler and compass point.*

Proof. To say that a finite sequence Q_1,Q_2,Q_3, ... ,Q_m of points satisfies the condition of Definition 2.1 means that for each of the Q_i we have $Q_i = (0,0)$, $Q_i = (1,0)$, or Q_i is obtained in one of the three ways denoted as (i), (ii), and (iii) in Definition 2.1. Suppose Z is a point of intersection of two ruler and compass lines, a point of intersection of a ruler and compass line and a ruler and compass circle, or a point of intersection of two ruler and compass circles. Now, (i) if point Z is a point of intersection of two ruler and compass lines, then there are ruler and compass points P,Q,R,S such that Z is a point in the intersection of $\overleftrightarrow{PQ}$ and $\overleftrightarrow{RS}$; (ii) if point Z is a point of intersection of a ruler and compass line and a ruler and compass circle, then there are ruler and compass points P,Q,R,S such that Z is a point in the intersection of $\overleftrightarrow{PQ}$ and R_S; and (iii) if point Z is a point of intersection of two ruler and compass circles, then there are ruler and compass points P,Q,R,S such that Z is a point in the intersection of P_Q and R_S. Then, in any of these three cases, there is a sequence P_1,P_2,... ,P; a sequence Q_1,Q_2,... ,Q; a sequence R_1,R_2,... ,R; and a sequence S_1,S_2,... ,S such that each of the four sequences satisfies the condition of Definition 2.1. So the sequence

$$P_1, P_2, \ldots, P, Q_1, Q_2, \ldots, Q, R_1, R_2, \ldots, R, S_1, S_2, \ldots, S$$

must satisfy the condition. Hence, in any of the three cases, the sequence

$$P_1, P_2, \ldots, P, Q_1, Q_2, \ldots, Q, R_1, R_2, \ldots, R, S_1, S_2, \ldots, S, Z$$

also satisfies the condition, and Z is therefore a ruler and compass point by Definition 2.1. ∎

Next, we want to validate our use of the modern compass in place of the euclidean compass.

Theorem 2.3. *If A,B,C are three ruler and compass points, then A_{BC} is a ruler and compass circle.*

Proof. Suppose A,B,C are three ruler and compass points. (See Figure 1.1; we follow Euclid's construction for I.2.) Let D be a point of intersection

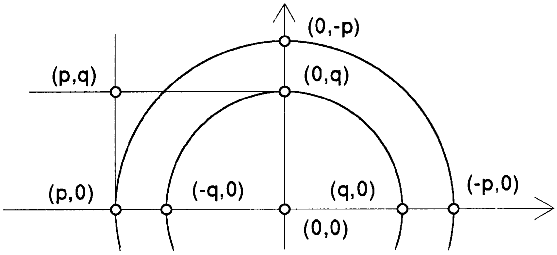


FIGURE 2.1.

of the two ruler and compass circles A_B and B_A. Then D is a ruler and compass point by the preceding theorem. Let E be the point of intersection of B_C and $\overrightarrow{DB}$ such that D–B–E. Then, since E is a point of intersection of the ruler and compass circle B_C and the ruler and compass line $\overleftrightarrow{DB}$, we have E is a ruler and compass point. Let F be the intersection of D_E and $\overrightarrow{DA}$. Since F is a point in the intersection of ruler and compass circle D_E and ruler and compass line $\overleftrightarrow{DA}$, then F is a ruler and compass point. Since $BC = AF$, then $A_{BC} = A_F$. Since A_F is a ruler and compass circle, then A_{BC} is a ruler and compass circle, as desired. ∎

After verifying the fundamental intersection properties of ruler and compass lines and ruler and compass circles in Theorem 2.2, it is hoped that you thought the proof of Theorem 2.3 was obvious from what we did in Chapter 1. We just repeated Euclid's construction with some new terminology. Indeed, that is the case, and we shall not continue to do this sort of thing. We will assume the basic constructions from the first chapter without further mention. For example, from what we know from Chapter 1, the next paragraph should be a convincing argument for the next theorem.

Of all the points (x, y) in the cartesian plane, we want to distinguish those that are the ruler and compass points. It will take some time to do this. We begin with the two points $(0, 0)$ and $(1, 0)$. So the X-axis is a ruler and compass line. Then $(-1, 0)$ must be a ruler and compass point, as this point is in the intersection of the X-axis and the circle through $(1, 0)$ with center $(0, 0)$. It soon follows that $(n, 0)$ is a ruler and compass point for any integer n. Since we can construct perpendiculars and parallels by the methods of Euclid I.11, I.12, and I.31, then the following theorem is evident after considering Figure 2.1.

Theorem 2.4. *The coordinate axes are ruler and compass lines. All of $(p, 0)$, $(-p, 0)$, $(0, p)$, and $(0, -p)$ are ruler and compass points if any one is a ruler and compass point. Number x is a ruler and compass number iff $-x$ is a ruler and compass number. The integers are ruler and compass numbers. Point (p, q) is a ruler and compass point iff both p and q are ruler and compass numbers.*

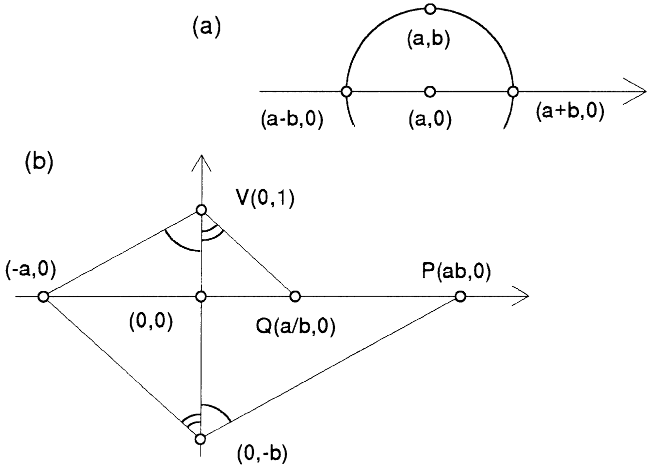


FIGURE 2.2.

Just because the X-axis is a ruler and compass line, it does not follow that every point on the X-axis is a ruler and compass point. Likewise, not every point on a ruler and compass circle is a ruler and compass point.

Recall that a rational number is a quotient m/n where m and n are integers with $n \neq 0$. A real number that is not a rational is called an irrational number. You may know a more general definition of the word "field" than that given below. However, such a generalization is not necessary for our purposes.

Definition 2.5. A ***field*** F is a subset of the real numbers that contains 0 and 1 and such that

$$a+b, \quad a-b, \quad ab, \quad a/c$$

are in F whenever a,b,c are in F but $c \neq 0$. Let $\mathbb{Q}$ denote the field of rational numbers, and let $\mathbb{R}$ denote the field of real numbers. Field F is ***euclidean*** if x in F and $x > 0$ implies $\sqrt{x}$ is in F.

Definition 2.5 assumes you know that the rationals form a field and that the reals form a field. The irrationals do not form a field. Some fields other than $\mathbb{Q}$ and $\mathbb{R}$ will be of considerable importance to us. The first of these appears in the next theorem, which follows from Figure 2.2. That is, after considering Figure 2.2, you should be able to provide the argument that proves the statement of Theorem 2.6. Such a proof hinges on being able to copy segments (Euclid I.3) and angles (Euclid I.23) with the ruler and compass.

Theorem 2.6. *The ruler and compass numbers form a field.*

The numbers in $\mathbb{Q}$ are ruler and compass numbers, but not every ruler and compass number is a rational. Not only can we perform the four arith-

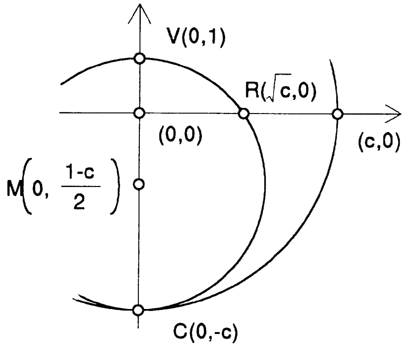


FIGURE 2.3.

metic operations of addition, subtraction, multiplication, and division on ruler and compass numbers, but we can also extract square roots. Hence, by Definition 2.5, we have a theorem that follows from considering Figure 2.3.

Theorem 2.7. *The ruler and compass numbers form a euclidean field.*

To say that x is a *square in field* F means that x is in F and there is a number y in F such that $y^2 = x$. In other words, number x in F is a square in F iff $\sqrt{x}$ is also in F. So, a field F is euclidean iff every positive number in F is a square in F. The field $\mathbb{Q}$ of rationals is not a euclidean field. For instance, 2 is certainly a positive rational but $\sqrt{2}$ is not a rational number. So 2 is a square in $\mathbb{R}$, but 2 is not a square in $\mathbb{Q}$. The field $\mathbb{R}$ of real numbers is a euclidean field, and very likely it is the only example of a euclidean field that you have seen before encountering the field of ruler and compass numbers.

We come now to an algebraic construction that is most important for our study of geometric constructions. The idea is to enlarge a given field to form a new field so that the new field contains some specific new number in addition to all the numbers of the given field. This turns out to be easy in the special case where the square of the specific number to be added is in the given field. For example, $\sqrt{5}$ is not a rational number but $(\sqrt{5}\,)^2$ is in $\mathbb{Q}$. We want to create the smallest field that contains $\sqrt{5}$ as well as all the rational numbers. With $F = \mathbb{Q}$ and $d = 5$ in the statement of the theorem below, the theorem says that the new field consists of all the numbers of the form $p + q\sqrt{5}$ with both p and q in $\mathbb{Q}$. It should be clear that the new field must contain these numbers at least; the theorem claims that the set of all these numbers does form a field. The theorem gives the general recipe for extending a given field by the square root of a positive number already in the field.

Theorem 2.8. *If F is a field and d is a positive number in F but $\sqrt{d}$ is not in F, then $\{p + q\sqrt{d} \mid p \text{ and } q \text{ in } F\}$ is a field.*

Because it is essential that the algebraic construction in this theorem be understood before going on, the proof of Theorem 2.8 is left as an exercise. To prove the theorem, it is necessary to show the sum, difference, product, and quotient of numbers $p_1 + q_1\sqrt{d}$ and $p_2 + q_2\sqrt{d}$ have the form $p_3 + q_3\sqrt{d}$ where the p_i and q_i are in F. In addition to proving the theorem, which you should do before reading the next definition, you should think of several examples. Consider an example where F is itself the field of all numbers of the form $p+q\sqrt{5}$ with p and q rational. This field will be denoted as $\mathbb{Q}(\sqrt{5}\,)$. Now it is not easy to determine whether an element in F is a square in F or not. For example, $(35 - 15\sqrt{5}\,)/2$ is a square in F since it is $[(5 - 3\sqrt{5}\,)/2]^2$. However, 2 is not a square in F since it can be shown that there are no rationals a and b such that $(a + b\sqrt{5}\,)^2 = 2$. With $d = 2$, the new field given by Theorem 2.8 is the set of all elements of the form $p + q\sqrt{2}$ where each of p and q is of the form $x + y\sqrt{5}$ with x and y in $\mathbb{Q}$. Hence, our new field has elements of the form $(p_1 + p_2\sqrt{5}\,) + (q_1 + q_2\sqrt{5}\,)\sqrt{2}$ where the p_i and q_i are rationals. If $F = \mathbb{Q}(\sqrt{5}\,)$, then this new field will be denoted as $F(\sqrt{2}\,)$ or $\mathbb{Q}(\sqrt{5}, \sqrt{2}\,)$. You can see that things get very complicated very fast; imagine how complicated it is to take $\mathbb{Q}(\sqrt{5}, \sqrt{2}\,)$ as our field F and play the game of Theorem 2.8 again. Fortunately, we deal with the theory and do not have to do much difficult computation in these complicated fields. The first sentence in the following definition formally introduces the notation $F(\sqrt{d}\,)$ for the field in Theorem 2.8.

Definition 2.9. If d is a positive number in field F but $\sqrt{d}$ is not in F, then $F(\sqrt{d}\,)$ denotes the field $\{p + q\sqrt{d} \mid p \text{ and } q \text{ in } F\}$. We call $F(\sqrt{d}\,)$ a ***quadratic extension*** of F. If $F_1 = F(\sqrt{d_1}\,)$, $F_2 = F_1(\sqrt{d_2}\,)$, ... , $F_n = F_{n-1}(\sqrt{d_n}\,)$, then we write $F_n = F(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_n}\,)$ and call each of F, F_1, F_2, ... , F_n an ***iterated quadratic extension of*** F. (The minor point that field F is defined to be an iterated quadratic extension of F but is not a quadratic extension of F is perhaps odd but turns out to be convenient.) Let $\mathbb{E}$ denote the union of all iterated quadratic extensions of the field $\mathbb{Q}$.

Once you get the idea of a quadratic extension from Theorem 2.8, then the idea of an iterated quadratic extension is just a quadratic extension of a quadratic extension of a quadratic extension ... of a quadratic extension. We allow only a finite chain of extensions. So $\mathbb{E}$ is the union of all iterated quadratic extensions of the rationals. The field $\mathbb{R}$ has no quadratic extensions, since every positive real is a square in $\mathbb{R}$. We are limiting ourselves to real numbers. A euclidean field has no quadratic extensions.

Figure 2.4 shows a *tower* of fields over the rationals. Field F_{i+1} is a quadratic extension of field F_i. So F_i is contained in F_{i+1}, and F_n is an iterated quadratic extension of the rationals. The union of the numbers in all such towers over the rationals is $\mathbb{E}$. Although we postpone answering the question of whether the numbers in $\mathbb{E}$ form a field, it is important to have a feeling for what numbers are in $\mathbb{E}$. What do they look like? They are

$$\begin{array}{l}
F_n = F_{n-1}(\sqrt{d_n}\,) = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n}\,) \\
\quad | \\
\quad \vdots \\
\quad | \\
F_{i+1} = F_i(\sqrt{d_{i+1}}\,) = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_{i+1}}\,) \\
\quad | \\
F_i = F_{i-1}(\sqrt{d_i}\,) = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_i}\,) \\
\quad | \\
\quad \vdots \\
\quad | \\
F_3 = F_2(\sqrt{d_3}\,) = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \sqrt{d_3}\,) \\
\quad | \\
F_2 = F_1(\sqrt{d_2}\,) = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}\,) \\
\quad | \\
F_1 = \mathbb{Q}(\sqrt{d_1}\,) \\
\quad | \\
\mathbb{Q}
\end{array}$$

FIGURE 2.4.

the real numbers you can write down using only parentheses, the integers, the signs $+$, $-$, $\times$, and $\div$ for the four arithmetic operations, and the sign $\sqrt{\ }$ for the operation of square root. An example is

$$\frac{915\sqrt{1 + 26546\sqrt{\sqrt{67}}} - \sqrt{5 + \sqrt{6 - \sqrt{7} + 5\sqrt{10 - 3\sqrt{\sqrt{6}}}}}}{\frac{314159}{100000}\sqrt{\sqrt{\sqrt{8 - \sqrt{57 - \sqrt{11}}}}}}$$

which you can punch into your calculator using no keys other than:

$$(\ )\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ +\ -\ \times\ \div\ \sqrt{\ }$$

If all the numbers in some field F are ruler and compass numbers, then the numbers in a quadratic extension $F(\sqrt{d}\,)$ are also ruler and compass numbers. This follows from Theorem 2.7. Therefore, by repeated application, if number x is in an iterated quadratic extension of the rationals, then x is a ruler and compass number.

Theorem 2.10. *If x is in $\mathbb{E}$, then x is a ruler and compass number.*

This last theorem should not be surprising, considering the form the numbers in $\mathbb{E}$ have. What we wish to do is prove the converse. This will take a bit of writing out and require four lemmas, but the ideas are not difficult. Think of starting with a set of points and then constructing new points with the ruler and compass. These new points are obtained as the intersection of lines and circles, and algebraically solving equations to find the coordinates of these points never requires anything more than the four arithmetic operations and taking square roots. So you would expect the coordinates of any point you can construct to be in an iterated quadratic extension of the rationals. Although you might consider this to be a convincing argument that proves Theorem 2.15, which is the converse of Theorem 2.10, the details are spelled out below. In any case, don't fail to see the principal argument by getting lost in the details.

Lemma 2.11. *If a line passes through two points each having coordinates in field F, then the line has an equation with coefficients in F. If both the center of a circle and a point on the circle have coordinates in field F, then the circle has an equation with coefficients in F.*

Proof. Since $(y_2 - y_1)X - (x_2 - x_1)Y + (x_2y_1 - x_1y_2) = 0$ is an equation for the line through the two points (x_1, y_1) and (x_2, y_2), the first part of the lemma depends only on the definition of a field.

Since equation $X^2 + Y^2 + (-2p)X + (-2q)Y + (s(2p-s) + t(2q-t)) = 0$ is equivalent to the equation $(X-p)^2 + (Y-q)^2 = (s-p)^2 + (t-q)^2$, which is an equation for the circle with center (p, q) that passes through (s, t), the second part of the lemma depends only on the definition of a field. ■

Lemma 2.12. *If each of two intersecting lines has an equation with coefficients in field F, then the point of intersection has coordinates in F.*

Proof. With $x_0 = (b_1c_2 - b_2c_1)/(a_1b_2 - a_2b_1)$ and $y_0 = (a_2c_1 - a_1c_2)/(a_1b_2 - a_2b_1)$, then (x_0, y_0) is the point of intersection of the line with equation $a_1X + b_1Y + c_1 = 0$ and the line with equation $a_2X + b_2Y + c_2 = 0$. Note that $a_1b_2 - a_2b_1 \neq 0$, as otherwise the lines would be parallel. Again, the lemma depends only on the definition of a field. ■

Lemma 2.13. *If a line and a circle intersect and each has an equation with coefficients in field F, then the points of intersection have coordinates in F or else in a quadratic extension of F.*

Proof. The line with equation $aX + bY + c = 0$ and the circle with equation $X^2 + Y^2 + fX + gY + h = 0$ intersect at the points (x_0, y_0) where

$$\begin{aligned} d &= (fb - ag)^2 + 4c(af + gb - c) - 4h(a^2 + b^2), \\ x_0 &= (abg - 2ac - b^2f \pm b\sqrt{d})/(2(a^2 + b^2)), \\ y_0 &= (abf - 2bc - a^2g \mp a\sqrt{d})/(2(a^2 + b^2)). \end{aligned}$$

We suppose a,b,c,f,g,h are in F. In order for the line and circle to intersect, d must be nonnegative. If d happens to be a square in F, then the coordinates of the points of intersection are in F. However, if d is not a square in F, then at least one of x_0 or y_0 is not in F since both a and b cannot be 0. In this case x_0 and y_0 are in $F(\sqrt{d})$, a quadratic extension of F. ∎

Lemma 2.14. *If each of two intersecting circles has an equation with coefficients in field F, then the points of intersection have coordinates in F or in a quadratic extension of F.*

Proof. The system

$$\begin{cases} X^2 + Y^2 + pX + qY + r = 0, \\ X^2 + Y^2 + fX + gY + h = 0 \end{cases}$$

of equations is seen to be equivalent to the system

$$\begin{cases} (p-f)X + (q-g)Y + (r-h) = 0, \\ X^2 + Y^2 + fX + gY + h = 0 \end{cases}$$

of equations by subtracting or adding the two equations in one system to get the first equation in the other system. Therefore, this lemma follows from the preceding lemma. ∎

We are now ready to state and prove the principal theorem. Following the proof, there is a lengthy example of the association of the points P_i with the fields F_i, as they appear in the proof.

Theorem 2.15. *The coordinates of a ruler and compass point are in an iterated quadratic extension of the field of rationals.*

Proof. Let P be a ruler and compass point. From the definition of a ruler and compass point, we see that P must be the last of a sequence P_1,P_2, ... ,P_n of points, each of which is $(0, 0)$, $(1, 0)$, or is obtained in one of three ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; (ii) as a point of intersection of a line through two points that appear earlier in the sequence and of a circle through an earlier point and having an earlier point as center; and (iii) as a point of intersection of two circles, each of which passes through an earlier point in the sequence and each of which has an earlier point as center. By the sequence of four lemmas, we can associate P_1 with the rationals and observe that each point P_i for $i > 1$ can be associated with a field F_i such that the coordinates of P_i are in F_i and such that F_i is either equal to F_{i-1} or else is a quadratic extension of F_{i-1}. Hence, F_n is an iterated quadratic extension of the rationals, and the coordinates of P lie in F_n. ∎

An example of the association of points and fields like that in the proof above has been promised. For this example, we construct the vertices of

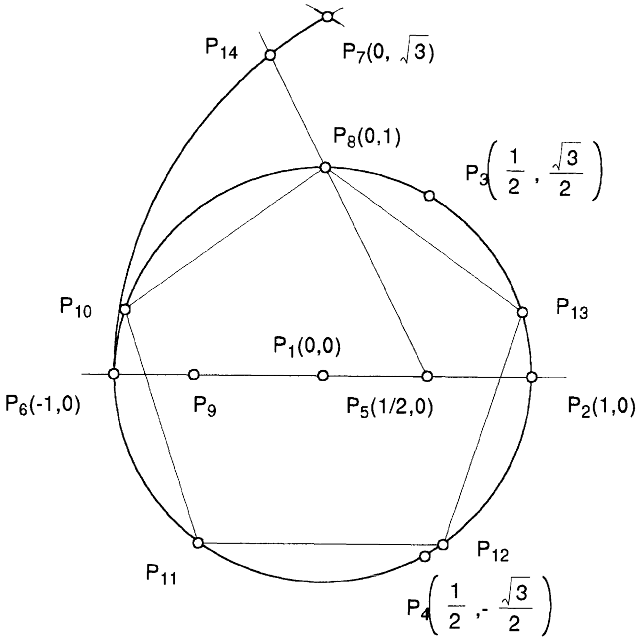


FIGURE 2.5a.

a pentagon inscribed in the unit circle. Figure 2.5a shows the vertices as P_8, P_{10}, P_{11}, P_{12}, and P_{13}. Then, only as a further algebraic illustration, P_{14} is thrown in for good luck. Starting with P_1 and P_2, we leave it to you to check that each point P_i is in fact a ruler and compass point by finding how the point is determined as an intersection of the proper type, as determined by the definition of a ruler and compass point and Theorem 2.2. For example, P_8 is the intersection of the line through P_1 and P_7 and of the circle through P_2 with center P_1. Figure 2.5b shows the field F_i associated with each point P_i. Of course, P_1 and P_2 are each associated with $\mathbb{Q}$. The coordinates of P_3 are in $\mathbb{Q}(\sqrt{3})$, a quadratic extension of F_2. Note that $\mathbb{Q}(\sqrt{3/4}) = \mathbb{Q}(\sqrt{3})$ and the latter notation is more concise for F_3. Further, although the coordinates of P_5 are in $\mathbb{Q}$, the field F_5 associated with P_5 in this sequence is still $\mathbb{Q}(\sqrt{3})$, since we want F_5 to be equal to F_4 or a quadratic extension of F_4. In general, we take F_k to be equal to F_{k-1} or a quadratic extension of F_{k-1}. Thus the coordinates of all points P_j with $j \leq k$ are in F_k. In order to introduce the $\sqrt{5}$ that appears in the coordinates of P_9, we need F_9 to be the quadratic extension of F_8 by $\sqrt{5}$. So $F_9 = F_8(\sqrt{5})$. A quadratic extension of F_9 is required for the coordinates of P_{10} to lie in F_{10}. We have $F_{10} = F_9(\sqrt{10+2\sqrt{5}})$. Since $\sqrt{10-2\sqrt{5}}\sqrt{10+2\sqrt{5}} = 4\sqrt{5}$, then the coordinates of P_{11} lie in F_{10}. So $F_{10} = F_{11} = F_{12} = F_{13}$. Then, as a final illustration, we need one more

$$\begin{array}{rcl}
((3-\sqrt{19})/5,\ (-1+2\sqrt{19})/5) = P_{14} & \leftrightarrow & F_{14} = \mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{10+2\sqrt{5}}, \sqrt{19}) \\
(\sqrt{10+2\sqrt{5}}/4,\ (-1+\sqrt{5})/4) = P_{13} & \leftrightarrow & F_{13} = \mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{10+2\sqrt{5}}) \\
(\sqrt{10-2\sqrt{5}}/4,\ (-1-\sqrt{5})/4) = P_{12} & \leftrightarrow & F_{12} = \mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{10+2\sqrt{5}}) \\
(-\sqrt{10-2\sqrt{5}}/4,\ (-1-\sqrt{5})/4) = P_{11} & \leftrightarrow & F_{11} = \mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{10+2\sqrt{5}}) \\
(-\sqrt{10+2\sqrt{5}}/4,\ (-1+\sqrt{5})/4) = P_{10} & \leftrightarrow & F_{10} = \mathbb{Q}(\sqrt{3}, \sqrt{5}, \sqrt{10+2\sqrt{5}}) \\
((1-\sqrt{5})/2,\ 0) = P_9 & \leftrightarrow & F_9 = \mathbb{Q}(\sqrt{3}, \sqrt{5}) \\
(0,\ 1) = P_8 & \leftrightarrow & F_8 = \mathbb{Q}(\sqrt{3}) \\
(0,\ \sqrt{3}) = P_7 & \leftrightarrow & F_7 = \mathbb{Q}(\sqrt{3}) \\
(-1,\ 0) = P_6 & \leftrightarrow & F_6 = \mathbb{Q}(\sqrt{3}) \\
(1/2,\ 0) = P_5 & \leftrightarrow & F_5 = \mathbb{Q}(\sqrt{3}) \\
(1/2,\ -\sqrt{3}/2) = P_4 & \leftrightarrow & F_4 = \mathbb{Q}(\sqrt{3}) \\
(1/2,\ \sqrt{3}/2) = P_3 & \leftrightarrow & F_3 = \mathbb{Q}(\sqrt{3}) \\
(1,\ 0) = P_2 & \leftrightarrow & F_2 = \mathbb{Q} \\
(0,\ 0) = P_1 & \leftrightarrow & F_1 = \mathbb{Q}
\end{array}$$

FIGURE 2.5b.

quadratic extension for the field F_{14}. We have $F_{14} = F_{13}(\sqrt{19})$. You should be convinced that no matter how many ruler and compass points are added to form the finite sequence of points P_i according to the rules set out in Definition 2.1, the lemmas assure us that the coordinates of P_i are in an iterated quadratic extension F_i of the rationals. This is the idea behind the proof of Theorem 2.15.

The theorem tells us that if x is a ruler and compass number then x must be in $\mathbb{E}$. This is the desired converse of Theorem 2.10. Note that it follows that $\mathbb{E}$ must be a field, since the ruler and compass numbers form a field. Since a field must contain 1 and hence all the rationals and since a euclidean field must be closed under square root, meaning that if x is in the field and $x > 0$ then $\sqrt{x}$ is in the field, we see that $\mathbb{E}$ must be the smallest euclidean field. As $\mathbb{Q}$ is the smallest field, so $\mathbb{E}$ is the smallest euclidean field.

Corollary 2.16. *Point P is a ruler and compass point iff the coordinates of P are in $\mathbb{E}$. Number x is a ruler and compass number iff x is in the field $\mathbb{E}$.*

We now have an algebraic handle on the ruler and compass points. Given any point in the cartesian plane, the point is constructible by means of ruler and compass iff the coordinates of the point are in an iterated quadratic extension of the rationals. We next determine some points that are not ruler and compass points. By finding points that cannot be constructed with ruler and compass, we will prove the impossibility of solving two of the three classical construction problems left unsolved by the Greeks:

(1) **TRISECTION OF THE ANGLE.** Given an angle, trisect the angle with ruler and compass.

(2) DUPLICATION OF THE CUBE. Given a segment of unit length, construct a segment of length $\sqrt[3]{2}$ with ruler and compass.

(3) SQUARING THE CIRCLE. Given a segment of unit length, construct a square having area π with ruler and compass.

Each of the three problems is unsolvable as stated. Note that this does not say there are no angles that can be trisected. Obviously, a right angle can be trisected with ruler and compass. To say the trisection problem is unsolvable is to say there are angles that can be constructed but cannot be trisected with ruler and compass. In order to prove the first two problems are unsolvable, we need to introduce some algebraic theorems.

Theorem 2.17. *If equation $ax^n + bx^{n-1} + \ldots + gx + h = 0$ has integer coefficients and has rational root p/q where p and q are integers having no common factor greater than* 1, *then p divides h and q divides a.*

Proof. Substituting p/q in place of x in the equation and multiplying the result through by q^n, we obtain $ap^n + bp^{n-1}q + \ldots + gpq^{n-1} + hq^n = 0$. Since p divides the right side of the equation and since p divides all the terms on the left side preceding the last, then p must divide hq^n. However, since p has no common factor greater than 1 in common with q, then p must divide h. Likewise, since q divides all the terms after the first but has no common factor greater than 1 in common with p, then q must divide a. ■

Theorem 2.18. *If a cubic equation with rational coefficients has no rational root, then none of its roots is a ruler and compass number.*

Proof. Assume a,b,c are rational numbers and the equation $x^3 + ax^2 + bx + c = 0$ has no rational root but does have a root that is a ruler and compass number. Let $F_0 = \mathbb{Q}$. Then there is a smallest positive integer k such that the cubic has a ruler and compass root r in an iterated quadratic extension F_k of the rationals with $F_k = \mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k})$. Let $F_i = \mathbb{Q}(\sqrt{d_1}, \ldots, \sqrt{d_i})$ for $i = 1,2, \ldots ,k - 1$. So F_i is a quadratic extension of F_{i-1} for $i = 1,2, \ldots ,k$. That is, $F_i = F_{i-1}(\sqrt{d_i})$. By our assumption, there exist p and q in F_{k-1} such that $r = p + q\sqrt{d_k}$. We must have $q \neq 0$, as otherwise r is in F_{k-1} and then k is not minimum. From the algebraic identity

$$\begin{aligned}(p \pm q\sqrt{d_k})^3 + a(p \pm q\sqrt{d_k})^2 + b(p \pm q\sqrt{d_k}) + c \\ = (p^3 + 3pq^2d_k + ap^2 + aq^2d_k + bp + c) \\ \pm (3p^2q + q^3dk + 2apq + bq)\sqrt{d_k},\end{aligned}$$

we see that $p - q\sqrt{d_k}$ must be a root of the cubic if $p + q\sqrt{d_k}$ is a root. The roots $p + q\sqrt{d_k}$ and $p - q\sqrt{d_k}$ are distinct since $q \neq 0$. Let t be the third root of the cubic. Then

$$\begin{aligned} &x^3 + ax^2 + bx + c \\ &\qquad = [x - t][x - (p + q\sqrt{d_k})][x - (p - q\sqrt{d_k})] \\ &\qquad\qquad = [x - t][x^2 - 2px + p^2 - q^2 d_k] \end{aligned}$$

for all x. By comparing the coefficients of x^2 from both sides, we must have $a = -t - 2p$. Hence, $t = -a - 2p$ and t is in F_{k-1}. However, this is a contradiction to the minimality of k, since t is a root of the cubic and is in F_{k-1}. Our initial assumption must be false. ∎

The following important lemma is left for Exercise 2.3.

Lemma 2.19. *There are three ruler and compass points P,Q,R such that $m\angle PQR = t$ iff $\cos t^\circ$ is a ruler and compass number.*

Recall the basic trigonometry formulas:

$$\sin(A + B) = \sin A \cos B + \cos A \sin B,$$
$$\cos(A + B) = \cos A \cos B - \sin A \sin B.$$

From these, we derive the following:

$$\sin 2A = 2 \sin A \cos A,$$
$$\cos 2A = (\cos A)^2 - (\sin A)^2 = 1 - 2(\sin A)^2 = 2(\cos A)^2 - 1,$$
$$\sin 3A = \sin(A + 2A) = [\sin A][4(\cos A)^2 - 1],$$
$$\cos 3A = \cos(A + 2A) = 4(\cos A)^3 - 3 \cos A.$$

Now, multiplying the last equation by 2 and rearranging the terms, we have the identity:

$$(2 \cos A)^3 - 3(2 \cos A) - 2 \cos 3A = 0.$$

Then, since $\cos[3(A + 120)]^\circ = \cos[3(A + 240)]^\circ = \cos[3A]^\circ$, we can easily check the following lemma.

Lemma 2.20. *Equation $x^3 - 3x - 2\cos(3A)^\circ = 0$ has roots*

$$2 \cos A^\circ, \qquad 2\cos(A + 120)^\circ, \qquad 2\cos(A + 240)^\circ.$$

Since $\cos 60^\circ = 1/2$, then by taking $3A = 60$ in Lemma 2.20 we see that the equation $x^3 - 3x - 1 = 0$ has roots $2\cos 20^\circ$, $2\cos 140^\circ$, and $2\cos 260^\circ$. Since, this equation has no rational root by Theorem 2.17, then none of these roots is a ruler and compass number by Theorem 2.18. In particular, we have proved the next very important theorem.

Theorem 2.21. *The number $\cos 20^\circ$ is not a ruler and compass number.*

By Lemma 2.19, the significance of the preceding theorem is that a 60° angle cannot be trisected with the ruler and compass. We have found an angle that cannot be trisected but that is (surprisingly?) easy to construct.

It follows that there are infinitely many such angles, but only one is sufficient to show the famous trisection problem has no solution. The problem called the TRISECTION OF THE ANGLE is unsolvable.

Theorem 2.22. *The number* $\sqrt[3]{2}$ *is not a ruler and compass number.*

The theorem above follows from Theorem 2.18 and that the equation $x^3 - 2 = 0$ has no rational root: The problem called the DUPLICATION OF THE CUBE is unsolvable.

As the ancient Greeks suspected, the three famous construction problems they left unsolved have no solutions. The first rigorous proof of the unsolvability of these first two problems was given by the little-known French mathematician Pierre Laurent Wantzel (1814–1848). The proof was published in 1837 when Wantzel was only twenty-three and an engineering student. Although frequently cited as P. Wantzel or P. L. Wantzel, his name appears on his published papers as L. Wantzel. His important 1837 paper has "Par M. L. Wantzel" under its title. This most likely accounts for the erroneous citations to M. L. Wantzel, where the French abbreviation M. for Monsieur has been confused as an initial. One thing is certain, L. Wantzel should be better known than he is.

The third classical construction problem left by the Greeks, SQUARING THE CIRCLE, is also unsolvable. To show this it is necessary and sufficient to show that π is not a ruler and compass number. From Exercise 2.14 we learn that every ruler and compass number is the root of a polynomial equation with integer coefficients. Such roots are called algebraic numbers and include the ruler and compass numbers. The real numbers that are not the roots of any polynomial equation with integer coefficients are called transcendental numbers. In 1882 the German mathematician Ferdinard Lindemann (1852–1939) showed that π is transcendental and thus proved that the last of the three classical construction problems also has no solution. That π is not an algebraic number requires a long proof, which we shall not give. Proofs may be found in the Dover paperback *Famous Problems of Elementary Geometry* by Felix Klein and in the nineteenth of the Carus Mathematical Monographs *Field Theory and its Classical Problems* by Charles Robert Hadlock and published by the Mathematical Association of America.

It is imperative to note the difference between an unsolved problem and an unsolvable problem. It is certainly reasonable to work on an unsolved problem in the hope of finding a solution by looking at the problem in a way different from anyone else. However, it is not reasonable to work on a problem that has been proved to be unsolvable. What we have shown above is that the problems called the Trisection of the Angle and the Duplication of the Cube cannot be solved. Not knowing if there is a solution and knowing there is not a solution are very different things.

It takes only a little more calculation than we have already done to show that it is impossible to construct a regular heptagon with the ruler and compass. We continue the development of the trigonometry formulas that was started above. We have:

$$\begin{aligned}\sin 4A &= \sin(2(2A)) = 2\sin 2A\cos 2A = 4(\sin A)(2\cos^3 A - 3\cos A),\\ \cos 4A &= \cos(2(2A)) = 2\cos^2 2A - 1 = 8\cos^4 A - 4\cos^2 A + 1,\\ \cos 7A &= \cos(3A + 4A) = 64\cos^7 A - 112\cos^5 A + 56\cos^3 A - 7\cos A.\end{aligned}$$

Multiplying the last equation by 2 and rearranging the terms, we have the identity:

$$(2\cos A)^7 - 7(2\cos A)^5 + 14(2\cos A)^3 - 7(2\cos A) - 2\cos 7A = 0.$$

From this it is easy to check that the equation

$$x^7 - 7x^5 + 14x^3 - 7x - 2 = 0$$

has the seven roots $x_k = 2\cos(360k/7)^\circ$ for $k = 0,1,2,3,4,5,6$. However these roots are not all distinct, since $x_1 = x_6$, $x_2 = x_5$, and $x_3 = x_4$, with $x_0 = 2$. You can check that the seventh degree equation factors as

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

and we have the first statement of the following theorem. It follows that a regular heptagon (7-gon) cannot be constructed with the ruler and compass. That it is impossible to construct a regular enneagon (9-gon) with the ruler and compass is a consequence of Theorem 2.21.

Theorem 2.23. *The equation* $x^3 + x^2 - 2x - 1 = 0$ *has roots* $2\cos(360/7)^\circ$, $2\cos(360 \times 2/7)^\circ$, *and* $2\cos(360 \times 3/7)^\circ$. *The numbers* $\cos(360/7)^\circ$ *and* $\cos(360/9)^\circ$ *are not in* $\mathbb{E}$.

Which regular polygons are constructible with the ruler and compass? On the one hand, the answer is known precisely; on the other hand, the final answer is that we do not know. To explain this, we must first learn about *fermat primes*, which are odd primes of the form $2^m + 1$. Pierre de Fermat (1601–1665) is perhaps best known for writing in a book that the margins were not large enough to contain his "marvelous demonstration" that $x^n + y^n = z^n$ has no solution in positive integers x,y,z,n with $n > 2$. Not finding among his papers this demonstration of what has come to be known as Fermat's Last Theorem, mathematicians have finally, after 300 years of effort, produced a proof.

Our concern is with Fermat's conjecture about odd primes of the form $2^m + 1$. Since

$$x^{2r+1} + 1 = (x+1)(x^{2r} - x^{2r-1} + x^{2r-2} - x^{2r-3} + \ldots - x^3 + x^2 - x + 1)$$

is an algebraic identity, letting $x = 2^q$ in the identity, we see that $2^{q(2r+1)} + 1$ is divisible by $2^q + 1$. So, if $2^m + 1$ is to have no nontrivial factors, then m

cannot have an odd factor greater than 1. So m must then be a power of 2. We take $m = 2^n$ and let

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

The F_n for $n = 0,1,2,3, \ldots$ are known as *fermat numbers.* Fermat was convinced by 1640 that all fermat numbers were prime and claimed in 1659 to have found the long-elusive demonstration for this conjecture. However, in 1732, Leonhard Euler (1701–1783), who ranks with Euclid, Archimedes, Newton, and Gauss among the greatest mathematicians, showed that F_5 is not a prime but the product of the two primes 641 and 6,700,417. Actually, no F_n for $n > 4$ is known to be prime. In 1992, F_{22} was shown to be composite. The number F_{22} has over a million digits. We also know that the gigantic number F_{23471} is not a prime. Today, as in Euler's time, there are only five known fermat primes:

n	0	1	2	3	4
F_n	3	5	17	257	65537

After finding a means of constructing the regular heptadecagon, the young genius of the highest order Carl Friedrich Gauss (1777–1855) went on to show that a regular p-gon is constructible for odd prime p if p is a fermat prime. The 1837 paper by the young Laurent Wantzel that first showed two of the ancient Greek construction problems are impossible also showed that if a regular p-gon is constructible for odd prime p, then p is a fermat prime. Gauss had stated that this was the case. Although he wrote that to try to prove otherwise was a waste of time, he gave no proof and no indication of a proof was found among his papers after his death. These results are put together with some minor lemmas to give the following complete result:

The Gauss–Wantzel Theorem. *A regular n-gon is constructible with ruler and compass iff n is an integer greater than* 2 *such that the greatest odd factor of n is either* 1 *or a product of distinct fermat primes.*

A formal statement in our theory would say that $\cos(360/n)^\circ$ is a ruler and compass number for integer n iff the greatest odd factor of n is 1 or a product of distinct fermat primes. Proof of the Gauss–Wantzel Theorem would take us too far astray. Proofs can be found in Hadlock's *Field Theory and its Classical problems*, mentioned above, and in *Ruler and the Round* by Nicholas D. Kazarinoff.

As has been said, there are only five known fermat primes at this time. It is possible that there are no more; it is possible that a new one will turn up tomorrow. On the one hand, the Gauss–Wantzel Theorem tells us precisely which regular polygons are constructible with the ruler and compass; on the other hand, we do not know the final answer until we know all the fermat primes.

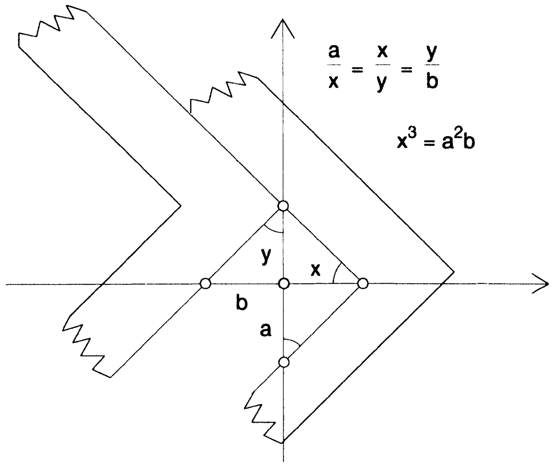


FIGURE 2.6.

Connected with the Gauss–Wantzel Theorem, there is one story whose telling cannot be avoided. From the theory, it is evident that a regular 65537-gon can be constructed with the ruler and compass. One person supposedly did so. Oswald Hermes spent ten years of his life carrying out the necessary constructions; the manuscript was deposited at the University of Göttingen in Germany. Nothing more need be said about Hermes.

The gateway to Plato's Academy proclaimed, Let no one ignorant of geometry enter here. Eudoxus was the greatest mathematician of the Academy and Menaechmus was his student. As tutor to the young man who would become Alexander the Great, Menaechmus informed his student that there was no royal road to geometry. (The same statement is said to have been told to King Ptolemy by Euclid.) The conic sections (the parabola, the ellipse, the hyperbola), which have become essential to the understanding of elementary science, were first studied at the Academy in connection with a theoretical problem of no practical application whatsoever. Menaechmus introduced conics to mathematics while working on the Delian Problem. In particular, Menaechmus used two parabolas for finding cube roots. (The simultaneous equations $x^2 = y$ and $y^2 = 2x$ of parabolas have the algebraic solutions $x = 0$ and $x = \sqrt[3]{2}$.) This is only one of the many times in history that mathematics developed without practical application in mind has later turned out to be exactly what was needed for some scientific advancement.

The finding of cube roots by using two carpenter's squares, as shown in Figure 2.6, is called Plato's Duplication of the Cube when $b = 2a$. By similar triangles in the figure, we have $a/x = x/y = y/b$, which will explain why the Delian Problem is also called the Problem of the Two Mean Proportionals. So $x^2 = ay$ and $xy = ab$. From these equations for

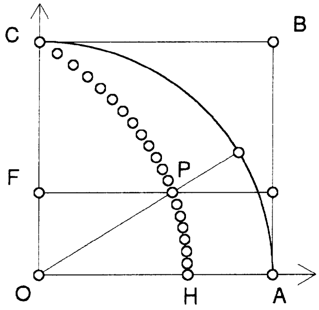


FIGURE 2.7.

a parabola and a hyperbola, we have $x^3 = xay = a^2b$. With $a = 1$, then $x = \sqrt[3]{b}$.

The Greeks resolved various problems by using cleverly designed intersections of conics. Today, with our use of cartesian geometry, these conic solutions are not exceedingly exciting. (Exercise 2.21 shows how Descartes used his coordinate geometry to give conic solutions to two classical problems. If you want to see how difficult conic sections are without cartesian geometry and why the great Greek mathematical machine was grinding to a halt under its own weight, start reading *Conics* by Apollonius of Perga.) The Greeks suspected that the unsolved classical constructions problems were, in fact, unsolvable. They lacked the algebra and analytic geometry needed to verify that ruler and compass solutions for these constructions are impossible.

Curves far more exotic than conics were invented in an attempt to solve the classical construction problems. The conchoid of Nicomedes is a mussel-shaped curve invented to trisect angles. Nicomedes also invented a device for tracing a conchoid. This tool is the oldest known construction tool other than the ruler and the compass. The cissoid of Diocles is an ivy-shaped curve that was invented to construct cube roots. The spiral of Archimedes was invented to square the circle. See Chapter VII of the first volume of the two volume classic *A History of Greek Mathematics* by Sir Thomas L. Heath. For other mechanical construction tools, see *The Trisection Problem* by Robert C. Yates, reprinted by The National Council of Teachers of Mathematics.

One of the exotic curves invented for angle trisection is the *quadratrix* of Hippias, who was born about 460 BC. Consider Figure 2.7 as you read the following description. Suppose $\square OABC$ is a square. Suppose a radius of O_A moves uniformly from $\overline{OC}$ to $\overline{OA}$ and at the same time the segment $\overline{CB}$ moves to $\overline{OA}$ uniformly and parallel to $\overline{OA}$. The two segments reach $\overline{OA}$ at the same time. The locus of the intersection of these two moving segments defines the quadratrix. Now, $(m\angle AOP)/(m\angle AOC) = OF/OC$ for point P on the quadratrix with F the foot of the perpendicular from P

to $\overline{OC}$. The curve changes all angle division problems to segment division problems. In particular, the angle trisection problem is reduced to segment trisection. For example, a line parallel to $\overline{OA}$ that trisects $\overline{OF}$ in Figure 2.7 will intersect the quadratrix at a point Q such that $\overrightarrow{OQ}$ trisects $\angle AOP$.

Menaechmus had a brother, Dinocratus, who was also a mathematician. Dinocratus discovered the most surprising property of the quadratrix. This property might be guessed from the name of the curve; after all, the curve is not known as the "trisectrix" of Hippias. What is the length of $\overline{OH}$? Taking $OA = 1$ and using radian measurement here, we see that the curve has the equation $y = x\tan(y\pi/2)$, $0 < y < 1$. The question about OH can be rephrased, What is the value of x as y approaches 0? Those with a knowledge of calculus can quickly show that $OH = 2/\pi$ in Figure 2.7, since

$$\lim_{y\to 0}\frac{y}{\tan(y\pi/2)} = \lim_{y\to 0}\frac{[2/\pi][\cos(y\pi/2)]}{[\sin(y\pi/2)]/[y\pi/2]} = 2/\pi.$$

Given O and H such that $OH = 2/\pi$, it is easy to construct a segment of length $\sqrt{\pi}$ and then to square a circle.

It did not escape the attention of the Greeks that there is a defect in the dynamical process defining the quadratrix. This defect disallows the use of the quadratrix as an ideal tool for squaring the circle. Our point H in Figure 2.7 is not obtained by the dynamic process. Each point of of the quadratrix is determined as the intersection of two segments. However, these segments coincide at the end of the process and, hence, do not intersect at a unique point. We can get as close as we like to H but we cannot reach H without resorting to a limit process. Approximations are exceedingly useful for many purposes, but they are not acceptable in our game.

The most famous of all geometric constructions that is not a ruler and compass construction is the trisection of the angle by Archimedes (287–211 BC). It is sufficient to be able to trisect acute angles since an angle of 30° is easily constructed. (The construction actually works for angles up to 145°.) The tool used by Archimedes is the marked ruler, which we will study in detail later. We suppose we have a ruler with two marks C and D placed one unit apart on the edge and that acute angle $\angle AOB$ is given with $OA = OB = 1$. The idea is to slide the marked ruler about until the ruler passes through B, has one mark on O_A, and has the second mark on $\overrightarrow{AO}$. See Figure 2.8 and suppose $m\angle OCD = x$. We want to show the measure of the given angle is $3x$. Now $m\angle COD = x$ by the *Pons Asinorum* (Euclid I.5: The base angles of an isosceles triangle are congruent). So $m\angle ODB = 2x$ by the *Exterior Angle Theorem* (Euclid I.32: The measure of an external angle of a triangle is the sum of the measures of the two opposite interior angles) applied to $\triangle ODC$. Then $m\angle OBD = 2x$ by the Pons Asinorum. Therefore, $m\angle AOB = 3x$ by the Exterior Angle Theorem applied to $\triangle OBC$.

As we have just seen, it is not difficult to trisect an angle. The Greeks certainly knew how to trisect angles with various tools. However, as we have

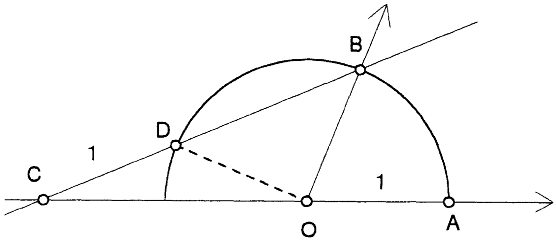


FIGURE 2.8.

proved, it is impossible to trisect an angle in general—and a 60° angle, in particular—using only straightedge and compass.

Exercises

2.1. If d is a positive element of field F but not a square in F and if $a + b\sqrt{d} = 0$ with a and b in F, then $a = b = 0$.◊

2.2. Prove Theorem 2.8.◊

2.3. Prove Lemma 2.19.◊

2.4. Prove there are ruler and compass points P,Q,R such that $m\angle PQR = t$ iff $\sin t°$ is a ruler and compass number; prove there are ruler and compass points P,Q,R such that $m\angle PQR = t$ iff $\tan t°$ is a ruler and compass number.

2.5. Show that $\sqrt{2}$ is irrational. In general, if d,m,n are positive integers but $d^{m/n}$ is not an integer, then show that $d^{m/n}$ is irrational.◊

2.6. Show that 2 is not a square in $\mathbb{Q}(\sqrt{5})$ and that $\sqrt{3}$ is not in $\mathbb{Q}(\sqrt{2})$.◊

2.7. Use only the theorems that we have proved to determine for which integers n an angle of $n°$ can be constructed with the ruler and compass. Also, verify your result using the Gauss–Wantzel Theorem.◊

2.8. For how many odd integers n are regular n-gons known to be constructible?◊

2.9. Illustrate Archimedes' trisection construction by constructing an angle of 20° with the marked ruler and compass.

2.10. For any field F, show that $\sqrt{d}$ is not a square in $F(\sqrt{d})$. (Corollary: An iterated quadratic extension of $\mathbb{Q}$ in not euclidean.)◊

2.11. Give an iterated quadratic extension of the rationals that contains $\sqrt{1 + \sqrt{2} - \sqrt{3} + \sqrt[4]{5 + \sqrt{6}}}$.

2.12. Find an example where an angle of $x°$ cannot be constructed with ruler and compass but a given angle of $(3x)°$ can be trisected with ruler and compass.◊

2.13. Find a polynomial equation with integer coefficients that has the number $\sqrt{3} - \sqrt[4]{2}$ as a root.◊

2.14. Prove the Lemma: If c is a root of a polynomial equation of degree k with coefficients in $F(\sqrt{d}\,)$, then c is a root of a polynomial equation of degree $2k$ with coefficients in F. Prove the Theorem: An element of $\mathbb{E}$ is the root of a polynomial equation with integer coefficients and whose degree is a power of 2.◊

2.15. Why do the numbers of the form $a + b\sqrt{2} + c\sqrt{3}$ with a,b,c rational not form a field? Do all the numbers of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$ with a,b,c rational form a field?◊

2.16. Find the roots of $3x^3 + 16x^2 + 22x + 8 = 0$.◊

2.17. If a,b,c are in $\mathbb{E}$ but not all 0 and if r is a real root of the equation $ax^2 + bx + c = 0$, then show that r is in $\mathbb{E}$.

2.18. Suppose d_1 and d_2 are positive numbers in field F but neither is a square in F. Show that $F(\sqrt{d_1}) = F(\sqrt{d_2})$ iff $\sqrt{d_1/d_2}$ is in F.◊

2.19. Do $\mathbb{Q}(\sqrt{2}, \sqrt{5}\,)$ and $\mathbb{Q}(\sqrt{5}, \sqrt{2}\,)$ contain the same elements of $\mathbb{R}$? Do $\mathbb{Q}(\sqrt{6}, \sqrt{2}\,)$ and $\mathbb{Q}(\sqrt{6}, \sqrt{3}\,)$ contain the same elements of $\mathbb{R}$? Show that $2(3 + \sqrt{5}\,)$ is a square in $\mathbb{Q}(\sqrt{5}\,)$ but that $10 + 2\sqrt{5}$ is not.◊

2.20. Which of the lemmas numbered 2.10 through 2.14 have a converse that is true?

2.21. Show how Descartes used the parabola with equation $y = x^2$ and the circle through $(0, 0)$ with center $(k/2, 1/2)$ to construct cube roots and used the parabola with equation $y = x^2$ and the circle through $(0, 0)$ with center $(\cos 3A, 2)$ to trisect angles.◊

2.22. Find out how the cissoid of Diocles can be used to construct cube roots, how the conchoid of Nicomedes can be used to trisect angles, and how the spiral of Archimedes can be used to square a circle.

2.23. Show that $\cos(360/17)°$ is in $\mathbb{E}$.◊

2.24. A Term Project: Prove the Gauss–Wantzel Theorem.

2.25. A Term Project: Prove π is not in $\mathbb{E}$.

$$2^{32} - 1 = 3 \times 5 \times 15 \times 257 \times 65537.$$

3

The Compass and the Mohr–Mascheroni Theorem

> In December of 1797 there took place in Paris a brilliant gathering of prominent writers and scholars, with the immortal Lagrange and Laplace among them. A most conspicuous member of the company was the young and victorious General Napoleon Bonaparte, who ... had occasion to entertain Lagrange and Laplace with a kind of solution of some elementary problems of elementary geometry that was completely unfamiliar to either of the two world-famous mathematicians. Legend has it that after having listened to the young man for a considerable while, Laplace, somewhat peeved, remarked, "General, we expected everything of you, except lessons in geometry."
>
> N. A. COURT

Napoleon proposed to the French mathematicians the problem of dividing a circle into four congruent arcs by using the compass alone. Although not original with Napoleon, the problem has become known as *Napoleon's Problem*. During his campaign in northern Italy, Napoleon had encountered the poet and geometer Lorenzo Mascheroni (1750–1800). Mascheroni

G. E. Martin, *Geometric Constructions*

was a professor at the University of Pavia, where Christopher Columbus had once been a student. Mascheroni's most famous mathematical work is his *Geometria del Compasso*, published in 1797. This work, which began with an ode of some literary merit that was dedicated to Napoleon, showed that all the ruler and compass constructions can be accomplished with the euclidean compass alone. Surprisingly, any point that can be constructed with ruler and compass can be constructed without using the ruler at all. In these compass constructions, a line is considered to be constructed as soon as two points on the line are constructed. In practice, we cannot draw a line with only a compass, but we may be able to construct some particular point on the line as the intersection of circles that are drawn with the compass. As usual, we do not expect every point on a constructed line to be constructible.

The Danish geometer Johannes Hjelmslev (1873–1950) quickly realized the importance of the copy of *Euclides Danicus* that his student had picked up at a second hand bookshop. Hjelmslev had the book republished in 1928. *Euclides Danicus* thus surfaced by accident some two and a half centuries after it had been first published in 1672. The author, Georg Mohr (1640–1697), had anticipated Mascheroni by 125 years. Although he was born in Copenhagen, Mohr left Denmark as a young man to live in Holland. The book appeared in 1672 in two editions, one in Danish and one in Dutch. The references that existed to *Euclides Danicus* before 1928 incorrectly assumed that the small book was a commentary on Euclid's *Elements* or only an edition of a part of Euclid's work.

Geometers Mohr and Mascheroni had in different countries and in different centuries independently demonstrated that one could dispense with the ruler in classical geometric constructions. This startling result of Mohr and Mascheroni is the principal theorem of this chapter.

Following the precedent of Definition 2.1, which defined "ruler and compass point," we want to define "a compass point" to be a point that can be constructed with the euclidean compass. Of all the points in the cartesian plane, we then want to distinguish those that are compass points. The two points $(0,0)$ and $(1,0)$ in the starter set from the ruler and compass constructions will again be sufficient to get us started. We have in mind the euclidean compass as our only tool. To form the following definition, take Definition 2.1 and cross out all the references to the ruler. The result should then contain no surprises.

Definition 3.1. In the cartesian plane, a point is a ***compass point*** if the point is the last of a finite sequence $P_1, P_2, P_3, \ldots, P_n$ of points such that each point is in $\{(0,0),(1,0)\}$ or is a point of intersection of two circles, each of which passes through an earlier point in the sequence and each of which has an earlier point as center. A ***compass line*** is a line that passes through two compass points. A ***compass circle*** is a circle through a compass point with a compass point as center. A number x is a ***compass number*** if $(x, 0)$ is a compass point.

The intersection properties of compass lines and compass circles are definitely not obvious. Analogous to Theorem 2.2, the points of intersection of two compass circles are quickly seen to be compass points. This time, however, it is by no means clear that the points of intersection of a compass line with either a compass circle or a compass line are compass points. This is what we must prove to show that a ruler and compass point is a compass point. That a compass point is a ruler and compass point is trivially true. You may find it astonishing that we will soon be able to prove that the intersection of two compass lines is a compass point.

Theorem 3.2. *A point of intersection of two compass circles is a compass point.*

Proof. If point Z is a point of intersection of two compass circles, then there are compass points P,Q,R,S such that Z is a point in the intersection of P_Q and R_S. Then there is a sequence $P_1, P_2, \ldots, P$; a sequence $Q_1, Q_2, \ldots, Q$; a sequence $R_1, R_2, \ldots, R$; and a sequence $S_1, S_2, \ldots, S$ such that each of the four sequences satisfies the condition of Definition 3.1. So the sequence

$$P_1, P_2, \ldots, P, Q_1, Q_2, \ldots, Q, R_1, R_2, \ldots, R, S_1, S_2, \ldots, S$$

must satisfy the condition. Hence, the sequence

$$P_1, P_2, \ldots, P, Q_1, Q_2, \ldots, Q, R_1, R_2, \ldots, R, S_1, S_2, \ldots, S, Z$$

also satisfies the condition, and Z is therefore a compass point by Definition 3.1. ∎

Theorem 3.3. *If P and Q are two compass points, then the perpendicular bisector of $\overline{PQ}$ is a compass line.*

Proof. The theorem follows immediately from the preceding theorem by considering the two points of intersection of P_Q and Q_P. Say these compass points are A and B. Then the compass line $\overleftrightarrow{AB}$ is the desired perpendicular bisector of $\overline{PQ}$. This follows from the theorem that the perpendicular bisector of a segment is the locus (set) of all points that are equidistant from the endpoints of the segment. ∎

The points P and Q in the proof above are the images of each other under the reflection in $\overleftrightarrow{AB}$; if the plane could be folded along $\overleftrightarrow{AB}$, then the points P and Q would fall together. The key idea behind most of our constructions in the Mohr–Mascheroni theory is that of reflections. Recall that the reflection in $\overleftrightarrow{AB}$ is defined as the mapping on the set of points in the plane that sends point P to itself if P lies on $\overleftrightarrow{AB}$ and otherwise sends P to P' where $\overleftrightarrow{AB}$ is the perpendicular bisector of $\overline{PP'}$. The image

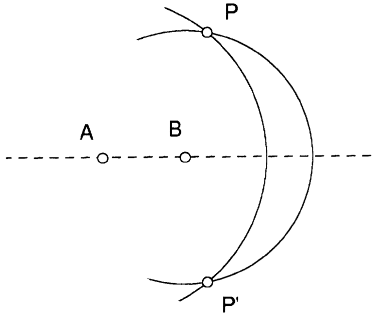


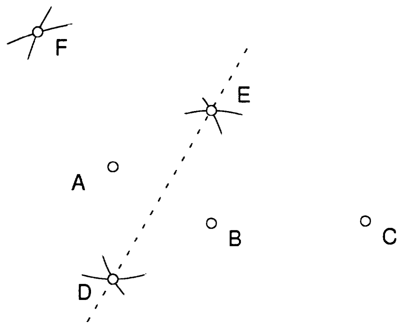
FIGURE 3.1.

P' of a point P under the reflection in $\overleftrightarrow{AB}$ is very easy to construct with a compass alone. We have $P = P'$ iff P is on $\overleftrightarrow{AB}$. Otherwise, P and P' are the two points of intersection of the circles A_P and B_P, as in Figure 3.1. Since $AP = AP'$ and $BP = BP'$ in this case, we have each of A and B is equidistant from P and P'. Thus $\overleftrightarrow{AB}$ is the perpendicular bisector of $\overline{PP'}$, and so the image of P under the reflection in $\overleftrightarrow{AB}$ is P' by the definition of a reflection. This simple construction is basic to many of the compass constructions.

Theorem 3.4. *The image of a compass point under the reflection in a compass line is a compass point.*

In making a sketch for a compass construction, beware of the trap of inadvertently assuming the intersection of two compass lines has been constructed. For example, the construction in Figure 3.1 does not provide us with the point of intersection of $\overleftrightarrow{AB}$ and $\overleftrightarrow{PP'}$. The point does exist. There is no question that there is such a point in the cartesian plane; the question is whether this point is a compass point or not. You may find it helpful to draw all segments and lines that you wish to emphasize with a colored pencil. Having the arcs and circles drawn in one color and the segments and lines drawn in another color may help you avoid the trap.

Euclid's proof of his proposition I.2 showed that the ruler and euclidean compass combination is equivalent to the ruler and modern compass combination. We shall now show that the euclidean compass and the modern compass are equivalent construction tools, without the presence of the ruler. Further, you may be very surprised that the construction for the next theorem, which states the equivalence of the euclidean compass and the modern compass, is much more simple than that given by Euclid for I.2. Actually, Euclid's construction and its proof are quite remarkable, considering that the construction is only the second proposition in the entire development of Euclid's geometry. The details of the proof of the more elegant construction that is indicated in Figure 3.2 are left to the reader. However, in the construction, $\overline{BC}$ is mapped to $\overline{AF}$ under the reflection in $\overleftrightarrow{DE}$. So $\overline{AF}$ and



A_B, B_A	D_C, E_C	A_F
D, E	C, F	

FIGURE 3.2.

$\overline{BC}$ are symmetric with respect to $\overleftrightarrow{DE}$ and, hence, are congruent. Therefore, if you know that segments are mapped to congruent segments under a reflection, then the proof is immediate. If you have to prove $AF = BC$ by chasing congruent triangles all over the place, then the proof is much longer.

Theorem 3.5 (*cf.* Euclid I.2). *If A, B, C are three compass points, then A_{BC} is a compass circle.*

There is a difference between the statement of Theorem 3.5 and the scheme in Figure 3.2. Theorem 3.5 merely states that it can be proved that A_{BC} is a compass circle. The theorem does not state that we actually know how to draw such a circle with a euclidean compass. The theorem is an existence theorem but not a construction; the theorem states something exists but does not imply that anyone knows how to find that something. The scheme in the figure, on the other hand, gives us the actual construction. Recall that a construction is a theorem that not only says something exists but gives an algorithm for finding that something. Thus the scheme stands in place of Theorem 3.6, which is stated below. As an indication of how to attack a proof of Theorem 3.5, we essentially gave the statement of Theorem 3.6 in the form of the scheme. A proof of Theorem 3.6 is a proof of Theorem 3.5, but not conversely.

Theorem 3.6 (*cf.* Euclid I.2). *If A, B, C are three compass points, points D and E are the points of intersection of A_B and B_A, and points C and F are the points of intersection of D_C and E_C, then A_{BC} is the compass circle A_F.*

Euclid I.3 is such an immediate consequence of Euclid I.2 in ruler and compass theory that it is difficult to think of I.2 without I.3. This is not the case here. Our Theorem 3.14, which is analogous to Euclid I.3 does not follow trivially from our Theorem 3.5, which is analogous to Euclid I.2. Theorem 3.5 says that we effectively have use of the modern compass. However, not even a modern compass allows us in one operation to "cut a segment off the end of a ray." To carry out that construction we will have to wait until we have shown that a compass circle and a compass line through the center of the circle intersect in compass points.

In spite of Theorem 3.5, compass constructions accomplished by using only the euclidean compass are deemed to have more style than those constructions that use the modern compass. That all the circles in a scheme have the form A_B and not the form A_{BC} indicates that the construction is carried out with a euclidean compass. A modern compass construction can be turned into a euclidean compass construction by means of Theorem 3.6. Of course, in making a construction drawing you can effectively make your modern compass into a euclidean compass by restricting its use. Although it is a matter of taste, the handicap of using only a euclidean compass does make the game of compass constructions more challenging.

Theorem 3.7 (The Compass Midpoint Theorem). *If A and B are compass points, M is the midpoint of A and B, and B is the midpoint of A and N, then M and N are compass points.*

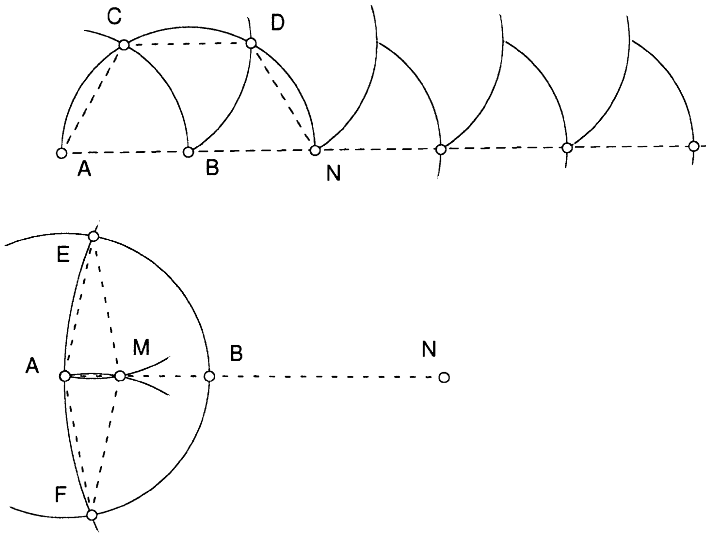
A construction for Theorem 3.7 is given by the scheme in Figure 3.3. There should be no problem with N; we encounter N halfway through the construction of a regular hexagon inscribed in B_A. The top of Figure 3.3 shows how to continue this process to get a segment of length nAB. A different, shorter construction for N is given in Exercise 3.3. The construction for the midpoint M does require some argument. Since $\triangle AEM \sim \triangle ANE$ because isosceles triangles with congruent base angles are similar, then

$$AM/AE = AE/AN = AB/(2AB) = 1/2$$

and $AB = AE = 2AM$, as desired. To divide a segment into n congruent parts, use the same idea as bisecting a segment, replacing $\overline{AN}$ by a segment of length $n(AB)$. The proof of the next theorem is left to the reader.

Theorem 3.8. *If A and B are two compass points and n is a positive integer, then points P and Q on $\overrightarrow{AB}$ such that $AP = nAB$ and $AQ = AB/n$ are compass points.*

We can now prove that the intersection of distinct compass lines is a compass point. In addition to using the preceding theorem in the construction, we use the following lemma: Three points P,Q,R are the vertices of an isosceles triangle if Q is the image of P under the reflection in $\overleftrightarrow{RS}$ but $\overline{PR} \not\perp \overline{RS}$. The lemma follows immediately from the definition of a reflection, since R is equidistant from P and Q if $\overleftrightarrow{RS}$ is the perpendicular bisector of $\overline{PQ}$.



A_B, B_A	C_A, B_A	D_C, B_A	N	A_B, N_A	E_A, F_A	M
C	A, D	C, N		E, F	A, M	

FIGURE 3.3.

Theorem 3.9. *The intersection of two compass lines is a compass point.*

Proof. Suppose A,B,C,D are four compass points with no three on one line such that $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ intersect at a unique point X. So $X \neq A$. Let E be the image of A under the reflection in $\overleftrightarrow{CD}$. If $\overleftrightarrow{AB} \perp \overleftrightarrow{CD}$, then X is the compass point that is the midpoint of the compass points A and E. Therefore, we suppose $\overleftrightarrow{AB} \not\perp \overleftrightarrow{CD}$. Let F be the compass point that is the image of E under the reflection in $\overleftrightarrow{AB}$. Let G be the compass point that is the image of A under the reflection in $\overleftrightarrow{EF}$. Then $\triangle AXE \sim \triangle AEG$, since both triangles are isosceles and have congruent base angles, and so $AX/AE = AE/AG$. We use this result below in the form $(AE)^2/AG = AX$. Possibly G_A and A_E do not intersect because $AE > 2AG$. Let n be a positive integer such that $AE < 2nAG$. (With luck, $n = 1$; in Figure 3.4, we have $n = 2$.) Then H on $\overrightarrow{AG}$ such that $AH = nAG$ is a compass point. So H_A and A_E do intersect at two compass points I and J. Next I_A and J_A intersect at two compass points A and K. Then $\triangle AIK \sim \triangle AHI$, since both triangles are isosceles and have congruent base angles, and so $AK/AI = AI/AH$. The point Y on $\overrightarrow{AK}$ such that $AY = nAK$ is a compass point. However,

$$AY = nAK = n[(AI)^2/AH] = n[(AE)^2/(nAG)] = (AE)^2/AG = AX.$$

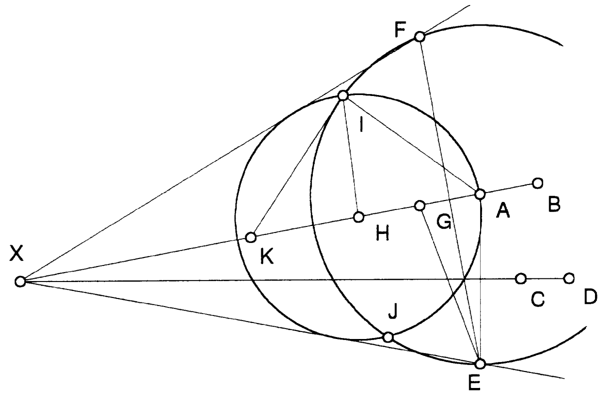


FIGURE 3.4.

Hence, since X and Y are on $\overrightarrow{AG}$, we have $X = Y$ and X is a compass point. ∎

In carrying out the construction given above for the intersection of two particular compass lines, it would be nice to be able to ignore that bothersome n. Suppose the given lines intersect in an acute angle of $t°$. Then $\sin t° = (AE/2)/AX = (1/2)(AE/AX) = (1/2)(AG/AE)$. So $AE < 2AG$ iff $\sin t° > 1/4$. This implies that if the acute angle between the lines is larger than 15° then you have the nice case where $n = 1$ and $G = H$ in the construction. Can you think how you might use this information in carrying out a construction? Think about it for a few minutes before looking at Exercise 3.8 to see how to formulate an equivalent construction problem to assure that the angle of intersection is larger than 15°.

We must show that the points of intersection of a compass line and a compass circle are compass points. This is extremely easy to do provided the line does not pass through the center of the circle. The idea behind the proof for this case is again that of a reflection. Since the points of intersection of a line and a circle also lie on the image of the circle under the reflection in the line, then the points are determined as the points of intersection of two circles. You should carry out the construction drawing to provide your own figure.

Theorem 3.10. *If A,B,C,D are compass points such that $\overleftrightarrow{AB}$ intersects C_D and if C is off $\overleftrightarrow{AB}$, then the points of the intersection of $\overleftrightarrow{AB}$ and C_D are compass points.*

Proof. Let C' and D' be the images of C and D, respectively, under the reflection in $\overleftrightarrow{AB}$. Then the compass circles C_D and C'_{D} intersect on $\overleftrightarrow{AB}$ at the desired points of intersection of $\overleftrightarrow{AB}$ and C_D. ∎

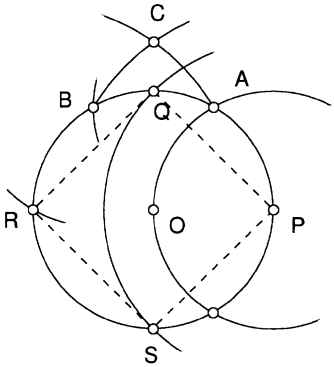


FIGURE 3.5.

Our goal of showing that we can accomplish the same constructions with the euclidean compass as with the ruler and compass now hinges on the possibility of constructing the points of intersection of a compass circle and a compass line through its center. What we might have expected to be a simple case in considering the intersection properties of compass lines and compass circles turns out to be the most complicated. We need to construct the intersection of $\overleftrightarrow{AB}$ and A_C for given compass points A,B,C. The special case where $\overleftrightarrow{AB} \perp \overleftrightarrow{AC}$ depends on a solution to Napoleon's Problem, which is to divide a circle into four congruent arcs using only a compass. It is easy to divide a circle into two semicircles; the hard part is to then bisect the semicircles. From another view, Napoleon's Problem asks for the construction of the vertices of a square inscribed in a given circle.

Theorem 3.11 (Napoleon's Problem, *cf.* Euclid IV.6). *If P and O are two compass points, then the points P,Q,R,S on O_P such that $\square PQRS$ is a square are compass points.*

Proof. The first construction given below in the form of a scheme is due to Mascheroni. This construction uses six new circles drawn with a modern compass and is easier to see than a second construction that is given later.

P_O, O_P	A_P, O_P	B_A, O_P	R_A, P_B	P_{OC}, O_P	P, Q, R, S
A	P, B	A, R	C	Q, S	

See Figure 3.5. In Mascheroni's construction, we take $OP = 1$ to simplify the algebra. Then $PR = 2$ and $PC = PB = s_3 = \sqrt{3}$. Since $\overline{OC}$ is perpendicular to $\overline{OP}$, then by the Pythagorean Theorem we have $PQ = PS = OC = \sqrt{PC^2 - OP^2} = \sqrt{2} = s_4$. ■

Although this proves the theorem, we go on to give a second construction due to W. F. Cheney in 1953 that uses only five new circles and furthermore these five circles are all drawn with a euclidean compass. An analysis of the

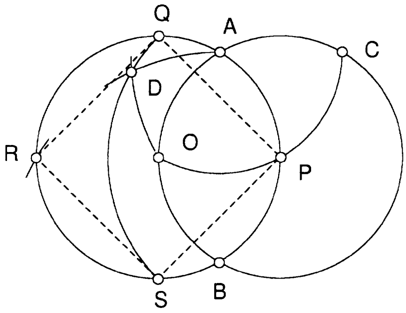


FIGURE 3.6.

first construction reveals that the proof depends on the theorem "2 = 3−1," restated in the form "$s_4^2 = s_3^2 - s_6^2$." If that construction has a blemish, it is that $\overline{OC}$ has neither of its endpoints on O_P. Cheney moves the same construction over to P_O and ends up with a segment of length s_4 with the endpoint P on O_P. In doing so, he needs to draw one less circle than Mascheroni, and all of these circles are drawn with a euclidean compass. Cheney's solution thus combines beauty with economy:

P_O, O_P	B_A, O_P	A_O, P_O	C_O, B_A	P_D, O_P	P, Q, R, S
A, B	A, R	O, C	D	Q, S	

See Figure 3.6. Here, still taking $OP = 1$, we have $PR = BC = 2$ and $CD = CO = s_3 = \sqrt{3}$. Since $\overline{PD}$ is perpendicular to $\overline{BC}$, then $PQ = PD = \sqrt{CD^2 - PC^2} = \sqrt{2} = s_4$, as desired.

Finding the intersection of a compass circle and a compass line through its center is still our concern. The problem is solved by a construction similar to Mascheroni's construction for Napoleon's Problem. The idea is to bisect certain arcs of a circle.

Theorem 3.12. *If A,B,C are three compass points, then the points of the intersection of $\overleftrightarrow{AB}$ and A_C are compass points.*

Proof. If C is on $\overleftrightarrow{AB}$ or if $\overleftrightarrow{AC} \perp \overleftrightarrow{AB}$, then the result follows from a solution to Napoleon's Problem. We suppose neither of these cases holds. By Theorem 3.2, then C', the image of C under the reflection in $\overleftrightarrow{AB}$, is a compass point different from C. The point D such that $\square ACC'D$ is a parallelogram is a compass point since D is a point of intersection of the compass circles $A_{CC'}$ and C'_A. See Figure 3.7. Likewise, D', the image of D under the reflection in $\overleftrightarrow{AB}$, is a compass point. Let E be a compass point on $\overleftrightarrow{AB}$ that is in the intersection of D_C and $D'_{C'}$. Let P and Q be the intersections we are looking for. We wish to show that P and Q are determined as the intersection of compass circles D_{AE} and D'_{AE}. That is, we want to show $DP = AE$. With F the foot of the perpendicular from C to

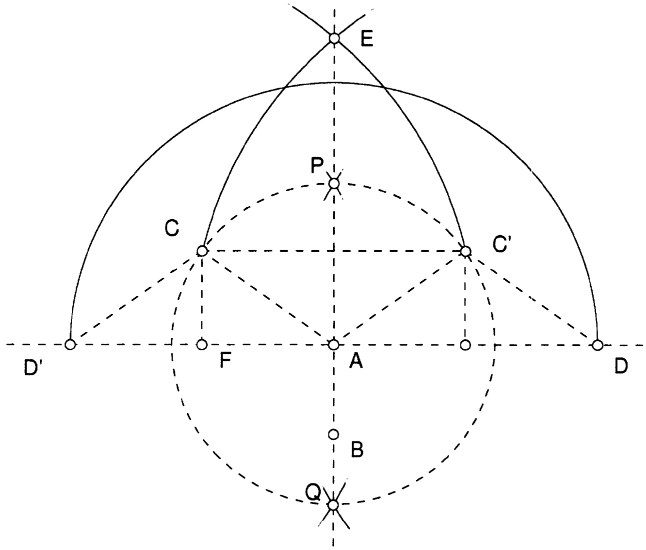


FIGURE 3.7.

$\overleftrightarrow{AD}$, then $AD = CC' = 2AF$ and $DF = 3AF$. Hence, by four applications of the Pythagorean Theorem, we have

$$\begin{aligned} DP^2 &= AP^2 + AD^2 = AC^2 + AD^2 \\ &= CF^2 + AF^2 + AD^2 \\ &= DC^2 - DF^2 + AF^2 + AD^2 = DC^2 - (3AF)^2 + AF^2 + (2AF)^2 \\ &= DC^2 - (2AF)^2 = DE^2 - AD^2 \\ &= AE^2, \end{aligned}$$

as desired. ∎

Putting Theorems 3.10 and 3.12 together, we have the following result.

Theorem 3.13. *The points of intersection of a compass line and a compass circle are compass points.*

Although the addition and subtraction of segments is not easy in compass constructions, that we effectively have the use of the dividers in compass constructions is a consequence of Theorem 3.5, which is analogous to Euclid I.2, and Theorem 3.13 above. This is stated as the first part of the next theorem; the second part is only a particular instance of the first part.

Theorem 3.14 (*cf.* Euclid I.3). *If A,B,C,D are compass points and if $A \neq B$, then the point E on $\overrightarrow{AB}$ such that $AE = CD$ is a compass point. If P and Q are compass points, then PQ is a compass number.*

Obviously, any point that can be constructed with a compass alone can be constructed with a ruler and compass. In other words, every compass point is a ruler and compass point. Comparing the definition of a ruler and compass point with the definition of a compass point, we see that we begin with the same starter set in both cases and that Theorems 3.2, 3.9, and 3.13 show no point is constructed with ruler and compass that cannot be constructed with the compass alone. Although it may be more work without the ruler, we obtain the same set of points in the end. In other words, every ruler and compass point is a compass point. We summarize our results in the following theorem.

Theorem 3.15 (The Mohr–Mascheroni Theorem). *A point is a compass point iff the point is a ruler and compass point.*

Although, we have just shown that any euclidean construction can be done with a compass alone, the theory does not tell us how to obtain nice compass constructions. The fun of seeking out elegant compass constructions remains. Each of Euclid's construction problems becomes a new problem to accomplish with the compass alone. For example, Euclid IV.11 asks for a regular pentagon inscribed in a circle. This problem is solved by each of the compass constructions of Exercises 3.12 and 3.13.

Until recently it was thought that the study of the rusty compass went back only as far as the Arab mathematician Abûl-Wefâ of Bagdad (940–998), who introduced the tangent function to trigonometry. As you have probably guessed, a *rusty compass* is a compass with a fixed opening; all circles drawn with a rusty compass have the same radius. A recent discovery of an Arabic translation of a work by Pappus of Alexandria, the last of the giants of Greek mathematics, shows that the study of the rusty compass has its roots in antiquity. In the history of geometry in the West, Pappus (*circa* AD 320) is followed by Descartes (*circa* 1637). However, compass constructions attracted the great German artist Albrecht Dürer (1471–1528) and many sixteenth-century Italian mathematicians, including Tartaglia, about whom we shall have more to say in Chapter 9. The Danish geometer Hjelmslev has considered using several rusty compasses. The Russian geometer Kostovskii has shown that restricting the compass so that the radii never exceed a prescribed length still leads to all compass constructible points, as does restricting the compass so that the radii always exceed a prescribed length. However, the problem of restricting the radii between a lower bound and an upper bound seems to be still open.

We will limit ourselves here to considering a rusty compass that draws unit circles. To model the rusty compass construction, we start with Definition 3.1, which defined a compass point. In this definition, we insert "rusty" before "compass," wherever it occurs, and we insert "unit" before "circle," wherever it occurs. Some pruning of a redundant phrase then gives Definition 3.16, below. This same process gives "a unit circle with a rusty

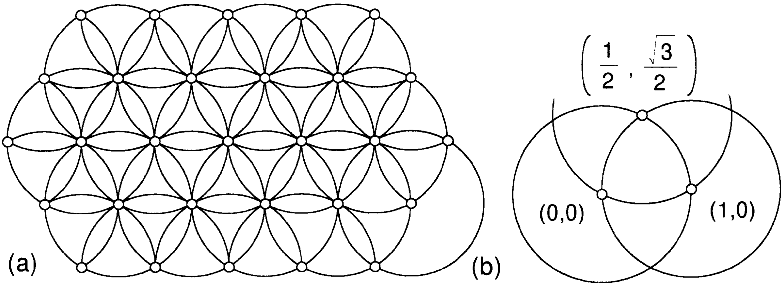


FIGURE 3.8.

compass point as center" as a definition for "a rusty compass circle." Perhaps "a unit circle through a rusty compass point and with a rusty compass point as center" would be better. Which do you think is the better definition? Remember definitions are made by people. Maybe we should take a vote. Until we have some compelling reason to prefer one of the suggested definitions over the other, we shall leave the matter undecided.

Definition 3.16. In the cartesian plane, a point is a ***rusty compass point*** if the point is the last of a finite sequence $P_1, P_2, P_3, \ldots, P_n$ of points such that $P_1 = (0,0)$, $P_2 = (1,0)$, and each of the others in turn is a point of intersection of two unit circles, each of which has an earlier point as center.

Is there anyone who has been alone with a compass and a piece of paper and who has not made a compass drawing like that in Figure 3.8a? Starting with unit circles with centers at $(0,0)$ and $(1,0)$ and drawing unit circles through the points of intersection already determined, we want to derive the coordinates of the points so constructed with the rusty compass. From Figure 3.8, where (b) requires knowing only the altitude of an equilateral triangle, it follows that we can get from $(0,0)$ to any rusty compass point by moving m times the directed distance from $(0,0)$ to $(1,0)$ and then n times the directed distance from $(0,0)$ to $(1/2, \sqrt{3}/2)$ where m and n are integers. The integers m and n may be zero, positive, or negative. Theorem 3.17 follows from this observation. Is it any wonder that the set of rusty compass points is called a *lattice*?

Theorem 3.17. *The rusty compass points are exactly the points*

$$(m + n/2, n\sqrt{3}/2)$$

where m and n are integers.

You may think that our starter set consisting of the points $P_1 = (0,0)$ and $P_2 = (1,0)$ is too nice. What happens if $P_1P_2 \neq 1$? We still suppose a unit circle is constructible if its center is constructible. Croft and Körner gave the following partial answer in 1978. If in Definition 3.16, we replace

"$P_2 = (1,0)$" by "$P_2 = (r,0)$" where $r < 2$, $r \neq 1$, and $r \neq 1/\sqrt{3}$, then there are constructible points arbitrarily close to any given point in the cartesian plane.

A vast number of problems are now open to you. As already mentioned, knowing that you can eventually do any ruler and compass construction problem with a compass alone is not at all the same thing as knowing a good compass construction for the problem. Suppose we give ourselves the ruler along with the rusty compass. What can be done now? Here is a whole new game. We can probably do more with ruler and rusty compass than with rusty compass alone. How much more? Is it possible to accomplish all ruler and compass constructions?

Exercises

3.1. Solve the compass construction problems that correspond to Euclid I.11 and I.12: Given compass points A and B, give a compass construction for a point E such that $\overline{AE} \perp \overline{AB}$. Given compass points A,B,F with F off $\overleftrightarrow{AB}$, give a compass construction for a point G such that $\overleftrightarrow{FG} \perp \overleftrightarrow{AB}$.$\diamondsuit$

3.2. Solve the compass construction problem that corresponds to Euclid I.31: Given compass points A,B,P with P off $\overleftrightarrow{AB}$, give a compass construction for a point Q such that $\overleftrightarrow{PQ} \parallel \overleftrightarrow{AB}$.$\diamondsuit$

3.3. Given points A and B, show that B is the midpoint of A and N:

A_B, B_A	D_C, C_D	N
C, D	N A–B–N	

3.4. Given compass points A and B, give a compass construction for a point D such that $m\angle ABD = 30$.$\diamondsuit$

3.5. Given compass points that are the vertices of an equilateral triangle, give a compass construction for the center of the triangle. Also, make a construction drawing to illustrate the construction.$\diamondsuit$

3.6. If the vertices of $\triangle ABC$ are compass points, then which of the following are impossible for $m\angle ABC$:

$$1, 2, 3, 4, 5, 10, 12, 15, 16, 18, 25, 32, 36, 38, 42?\diamondsuit$$

3.7. Is the set $\{(m + n/2, n\sqrt{3}/2) \mid m \text{ and } n \text{ integers}\}$ of points the same as the set $\{(s/2, t\sqrt{3}/2) \mid s \text{ and } t \text{ integers}\}$? Show the intersection of two rusty compass lines is not necessarily a rusty compass point. Show the midpoint of two rusty compass points is not necessarily a rusty compass point.$\diamondsuit$

3.8. Double an acute angle: Given compass points A,B,C,D such that $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ intersect at point X in acute angle $\angle AXC$, find compass line $\overleftrightarrow{EF}$ that is concurrent with $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ and such that $m\angle AXE = 2m\angle AXC$.◊

3.9. Make a construction drawing for the construction illustrated in Figure 3.7.

3.10. Make a construction drawing that illustrates Theorem 3.9.

3.11. Make a construction drawing for Mascheroni's construction and a construction drawing for Cheney's construction, where the constructions solve Napoleon's Problem (Theorem 3.11).

3.12. Make a construction drawing and prove Mascheroni's construction, which is given below, for inscribing a pentagon in a circle, where the construction continues from his solution to Napoleon's Problem, with $OP = 1$.◊

Q_O, O_P	E_{OC}, D_{OC}	$s_{10} = OF$ and $s_5 = FP$
D, E	F Q–O–F	

3.13. Make a construction drawing and prove Cheney's construction, which is given below, for inscribing a pentagon in a circle, where the construction continues from his construction to Napoleon's Problem, with $OP = 1$ and Q inside A_O.◊

Q_O, O_P	A_O, O_P	G_E, P_D	$s_{10} = OH$ and $s_5 = FH$
E, F E inside P_O	P, G	H H inside O_P	

3.14. Given noncollinear compass points A,B,C, then outline a compass construction for a point P such that $\overrightarrow{BP}$ bisects $\angle ABC$.

3.15. Given the endpoints of segments of lengths p,q,r, then outline a compass construction for a segment of length x where $p/q = r/x$.◊

3.16. Beginning with only compass points $A = (-1, 0)$, $B = (d, 0)$ with $d > 0$ but $d \neq 1$, $O = (0, 0)$, and $I = (0, 1)$, give a compass construction for $P = (\sqrt{d}, 0)$.◊

3.17. Suppose compass points A,B,C are such that A is on the perpendicular bisector of $\overline{BC}$. Give a compass construction for the center of the circle through A,B,C.◊

3.18. Give a compass construction for the circle through three given compass points.◊

3.19. Give a compass construction for a regular 30-gon that requires only eleven circles of four different radii. You may have to look this one up in the January 1944 *American Mathematical Monthly* (pp 47–48, Vol 51, Problem E567).

3.20. With one rusty compass, draw two circles having radii of different lengths. (This is a mind buster that the author assigns for April 1, even if class does not meet that day. The strange thing is that the problem can be done.)◊

4
The Ruler

> **line** (līn) *n.* **1. a.** The locus of a point having one degree of freedom; curve. **b.** A set of points (x, y) that satisfy the linear equation $ax + by + c = 0$, where a and b are not both zero. **2. a.** A thin, continuous mark, as that made by a pen, pencil, or brush applied to a surface. **b.** A similar mark cut or scratched into a surface. **c.** A crease in the skin, esp. on the face; a wrinkle. ... **32.** *Football.* **a.** A line of scrimmage. **b.** The linemen ... [ME<OE, cord, and OFr. *ligne*, line, both<Lat. *linea*, string, *linum* <thread.]
>
> *American Heritage Dictionary*
> Second College Edition

A culture developed in such an entangled jungle that the culture has no motivation or use for the concept of a straightedge is necessarily limited. The ruler and the compass are such basic tools that most of us accept them without ever having considered the implication of life without them. Suppose you are faced with the problem of coming to an examination on geometric constructions only to find that in making sure that you have not forgotten your compass you have completely forgotten to bring a ruler. How can you quickly make a straightedge to get you through the examination?

G. E. Martin, *Geometric Constructions*

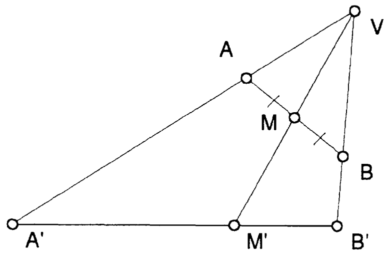


FIGURE 4.1.

Ripping your clothes apart to get a thread to pull taut is of theoretical interest but is not the most practical approach. Although reflections will not play the major role here that they did in the last chapter, the idea of a reflection does provide the construction of a straightedge. You have only to fold a sheet of paper; the crease forms a straightedge.

We have seen that the ruler and compass constructions can be associated with the field $\mathbb{E}$. The constructions that can be accomplished with the compass alone are also associated with $\mathbb{E}$, since the points that can be constructed are the same in either case. Can there be any doubt about what we should look at next? We want to know what constructions can be done with the ruler, without the compass. Is there a number field associated with the ruler constructions?

There is a mistaken impression that there are no worthwhile constructions that can be accomplished with the ruler alone. This is probably because of the following type of discussion, which argues that even a ruler construction for finding the midpoint of two given points cannot exist. Assume, to the contrary, that there is such a construction. Then, since any construction must be finite, all the relevant points necessary for the construction must fit inside some large circle. Here we invoke the third dimension and incline the plane of this circle to a second plane, with all of the circle on one side of the second plane. From a point V as in Figure 4.1, project the construction for bisecting $\overline{AB}$ from the first plane to the second plane so that any point P in the circle in the first plane corresponds to a point P' in the second plane such that V,P,P' are collinear. Figure 4.1 shows only the plane section determined by V, A, and B. Lines of the construction are projected onto lines of the second plane. Think of V as a light source and that the construction drawn on the first plane casts a shadow onto the second plane. So the assumed construction for the midpoint of segment $\overline{AB}$ in the first plane must go to a ruler construction for the midpoint of its image $\overline{A'B'}$, in the second plane. That is, the algorithm for obtaining the midpoint by drawing only lines in the first plane is followed in the second plane. However, there is a fly in the ointment. Midpoints of segments just do not generally get carried to midpoints of segments under the projection described. The method fails in the second plane. In Figure 4.1, the algo-

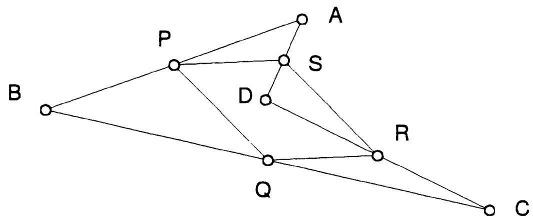
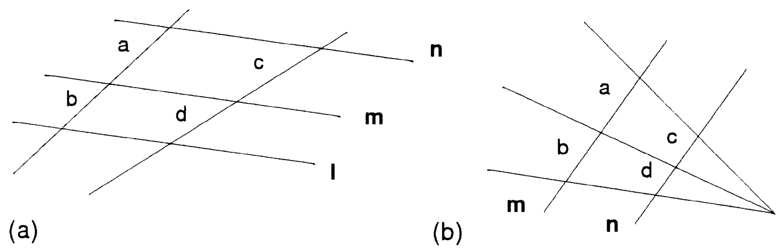


FIGURE 4.2.

rithm gives M' and not the midpoint of $\overline{A'B'}$; the shadow construction does not work. The argument concludes that the assumed method in the first plane must also have been bogus. The reason for then concluding that there are no worthwhile ruler constructions is based on the misunderstanding of the initial conditions allowed for ruler constructions. The properties of a starter set that is used as a reference may not be preserved under the projection. Hence, a construction may work in one plane, with the shadow construction failing in its plane because the shadow of the starter set may lack the properties of the starter set.

A similar consideration argues that a ruler construction of a square or even of two perpendicular lines is impossible. You can hold a regular sheet of paper in the sunshine so that the right angle from one corner casts a shadow of any acute angle or of any obtuse angle. Thus, the shadow construction for a right angle in the plane of the floor that follows a ruler construction for a right angle in the plane of the paper necessarily fails. Without this failure of preserving angle measure under parallel projection, sundials would be impossible; the sun's rays are considered parallel for most practical purposes. While performing this shadow experiment, notice that the shadow of the rectangular sheet of paper is always a parallelogram and that a midpoint of a side of the rectangle casts a shadow that is the midpoint of the corresponding side of the shadow parallelogram. Parallel projection preserves parallels and bisected segments. We will have a starter set that immediately implies a known right angle and some known bisected segments.

Suppose we are faced with a large sheet of paper and a ruler. We have an idea of what a ruler does. As usual, we leave nothing to chance and so must have a starter set in order to put down the ruler at all. It is clear that if only three noncollinear points are given, then we are at the end of all possibilities after drawing three lines. So we need at least the vertices of a quadrilateral to get off the ground. Then, if we can find the midpoints of the sides of this quadrilateral, we may as well suppose we are given the vertices of a parallelogram in our starter set. This follows from *Varignon's Theorem*: If P,Q,R,S are the midpoints of sides $\overline{AB}, \overline{BC}, \overline{CD}, \overline{DA}$ of any quadrilateral $\square ABCD$, then $\square PQRS$ is a parallelogram. See Figure 4.2; the proof of Varignon's Theorem is left for Exercise 4.14. However, given



a/c = b/d = (a + b)/(c + d)

FIGURE 4.3.

only the vertices of some parallelogram, we can then draw only six lines, first the four lines through the sides and then the two diagonals. We are then at a standstill. We want a theory that involves more than six lines. What can be done? One possibility is a five-point starter set. However, we will start with the four vertices $(1,0)$, $(0,1)$, $(2,0)$, and $(0,2)$ of a trapezoid. After some development, it will be seen with hindsight that this rather arbitrary choice for a starter set has been sufficiently generous. We are determined there be a construction theory for the ruler alone, and to accomplish this we have picked what turns out to be a convenient starter set.

With the starter set we have chosen, it is immediately obvious that $(0,0)$ will be constructible. Hence, you might say that we are effectively starting with the vertices of an isosceles right triangle and the midpoints of its legs. It is more traditional to describe ruler constructions as "constructions with ruler and given square." Accept for the moment that our starter set will give both $(0,0)$ and $(1,1)$ as constructible points, and so we have the vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ of a square. The "given square" is meant to imply that any intersection of the square and a constructed line is to be considered a constructed point. Bisected segments are provided by the diagonals of the square. However, as was pointed out above, starting with the vertices of a square leaves us with only six lines unless we have another point given. Even so, "constructions with ruler and given square" does describe the spirit of our theory. The vertices of a square and the midpoint of one side of the square would be a nice starter set. You may have correctly concluded that midpoints and parallels are important in the study of ruler constructions. You may be surprised to learn that they are so closely related to each other that you can't have one without the other, as we shall soon see.

We suppose familiarity with the elementary theorems on proportion that deal with segments and parallel lines; these are illustrated in Figure 4.3, where a,b,c,d are lengths of segments and lines l,m,n are parallel. Before turning to the definitions for ruler constructions, we interject a theorem from elementary geometry that may not be familiar but is very useful here.

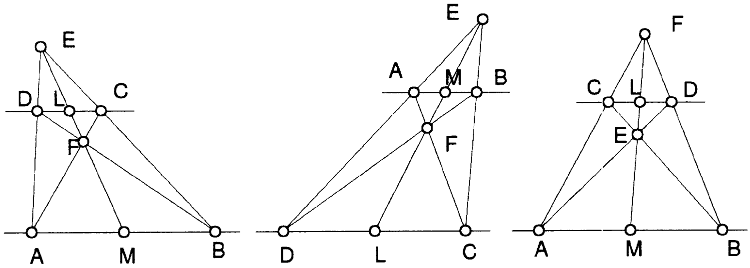


FIGURE 4.4.

Theorem 4.1 (The Trapezoid Theorem). *If $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ are distinct parallel lines with $\overleftrightarrow{AD}$ intersecting $\overleftrightarrow{BC}$ at E and with $\overleftrightarrow{AC}$ intersecting $\overleftrightarrow{BD}$ at F, then $\overleftrightarrow{EF}$ bisects $\overline{AB}$ and $\overline{CD}$.*

Proof. There are three cases: A–D–E, D–A–E, and A–E–D. See Figure 4.4. In any case, let $\overleftrightarrow{EF}$ intersect $\overline{AB}$ at M and $\overline{CD}$ at L. Then $\triangle AME \sim \triangle DLE$ and $\triangle BME \sim \triangle CLE$. So $AM/DL = ME/LE$ and $ME/LE = BM/CL$. Thus $AM/BM = DL/CL$. Also, $\triangle AMF \sim \triangle CLF$ and $\triangle BMF \sim \triangle DLF$. So $AM/CL = MF/LF$ and $MF/LF = BM/DL$. Thus $AM/BM = CL/DL$. Since $DL/CL = AM/BM = CL/DL$, then $DL = CL$ and $AM = BM$. ∎

Corollary 4.2. *If M is the midpoint of $\overline{QR}$, point P is off $\overleftrightarrow{QR}$ with S–P–Q, segments $\overline{PR}$ and $\overline{SM}$ intersect at T, and $\overleftrightarrow{QT}$ intersects $\overline{RS}$ at U, then $\overleftrightarrow{PU} \parallel \overleftrightarrow{QR}$.*

Proof. By the Trapezoid Theorem, if the parallel to $\overleftrightarrow{QR}$ that passes through P intersects $\overline{SR}$ at U', then $\overline{QU'}$, $\overline{PR}$, and $\overline{SM}$ are concurrent. See Figure 4.5. Hence, T is on $\overrightarrow{QU'}$. So $U' = U$ and $\overleftrightarrow{PU} \parallel \overleftrightarrow{QR}$, as desired. ∎

If we start with parallel lines, then Theorem 4.1 gives a construction for bisecting segments on the lines. If we start with bisected segments on a line, then Corollary 4.2 gives a construction for lines parallel to the line.

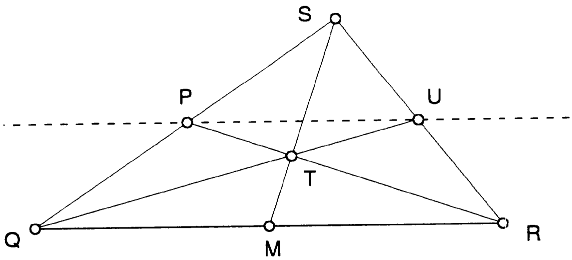


FIGURE 4.5.

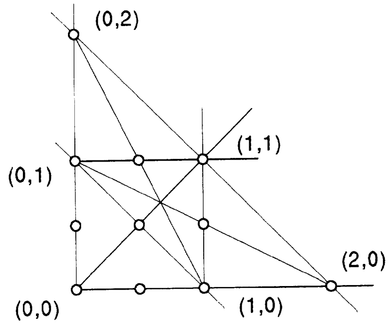


FIGURE 4.6.

This second construction augments the catalog of ruler and compass constructions for Euclid I.31 associated with Figure 1.12.

The definition below is formed by taking Definition 2.1, which defines a ruler and compass point, and deleting all the references to the compass. In the same way, the proof of the next theorem is derived from the proof of Theorem 2.2.

Definition 4.3. In the cartesian plane, a point is a ***ruler point*** if the point is the last of a finite sequence $P_1,P_2,P_3, \ldots ,P_n$ of points such that each point is in $\{(1,0),\ (0,1),\ (2,0),\ (0,2)\}$ or is the intersection of two lines, each of which passes through two points that appear earlier in the sequence. A ***ruler line*** is a line that passes through two ruler points. A ***ruler circle*** is a circle through a ruler point with a ruler point as center. A number x is a ***ruler number*** if $(x, 0)$ is a ruler point.

Theorem 4.4. *The intersection of two ruler lines is a ruler point.*

Proof. Suppose Z is a point of intersection of two ruler lines. Then there are ruler points P,Q,R,S such that Z is a point in the intersection of $\overleftrightarrow{PQ}$ and $\overleftrightarrow{RS}$. Further, there is a sequence $P_1,P_2, \ldots ,P$; a sequence $Q_1,Q_2, \ldots ,Q$; a sequence $R_1,R_2, \ldots ,R$; and a sequence $S_1,S_2, \ldots ,S$ such that each of the four sequences satisfies the condition of Definition 4.3. So the sequence

$$P_1, P_2, \ldots , P, Q_1, Q_2, \ldots , Q, R_1, R_2, \ldots , R, S_1, S_2 \ldots , S$$

must satisfy the condition. Hence, the sequence

$$P_1, P_2, \ldots , P, Q_1, Q_2, \ldots , Q, R_1, R_2, \ldots , R, S_1, S_2 \ldots , S, Z$$

also satisfies the condition, and Z is therefore a ruler point by Definition 4.3. ∎

We need to know that there are some ruler points around. In particular, we shall want to know that there is a third ruler point collinear with any two ruler points. The starter set and Theorem 4.4 provide us with the following

FIGURE 4.7.

lemma. You should duplicate Figure 4.6 by starting with the starter set and adding additional points as you follow the proof of the lemma.

Lemma 4.5. *Points* $(0,0)$,$(1,0)$,$(1,1)$,$(0,1)$ *are ruler points and the vertices of a square. The midpoints of the sides of this square are ruler points. Every ruler line contains three ruler points such that one is the midpoint of the other two.*

Proof. The six lines determined by pairs of points in $\{(1,0),\ (0,1),\ (2,0),\ (0,2)\}$ have equations $X = 0$, $Y = 0$, $X + Y = 1$, $X + Y = 2$, $2X + Y = 2$, and $X+2Y = 2$. Hence $(0,0)$ and $(2/3, 2/3)$ are ruler points. Then $Y = X$ is a ruler line. So $(1,1)$ and $(1/2, 1/2)$ are ruler points. Then $X = 1$ and $Y = 1$ are ruler lines. So $(1/2, 1)$ and $(1, 1/2)$ are ruler points. Then $X = 1/2$ and $Y = 1/2$ are ruler lines. So $(1/2, 0)$ and $(0, 1/2)$ are ruler points. For the last part of the lemma, any line that does not intersect each of the parallel ruler lines $X = 0$, $X = 1/2$, and $X = 1$ must intersect each of the parallel ruler lines $Y = 0$, $Y = 1/2$, and $Y = 1$. In either case, the three points of intersection determine a bisected segment. ∎

It is confession time. When we need a point either off or on some constructed line in a proof of a construction, we can easily verify its existence by citing Lemma 4.5. In drawing a construction, we have likewise been meticulous in insisting that we are not allowed to pick an arbitrary point. Each point must be constructed by the rules. Yet, to someone looking over our shoulder it seems otherwise. This is because this uninvited intruder does not see that, at least in our mind's eye, we have the figure in Figure 4.6 as part of our construction. We know that it is "there," even when it is not there. This can be dangerous if not exercised with caution. We should stop to rationalize the mental construction of our apparently picking an arbitrary point, even though we do not actually execute the formal construction. We are not being dishonest when we do this; we are being justifiably lazy.

As a consequence of the Trapezoid Theorem and its corollary we have a theorem that states a certain equivalence suggested by Figure 4.7. Note the figure is just a picture that may help you associate these two important concepts, namely midpoints and parallels. The details of this association are in the next theorem, which only restates the Trapezoid Theorem and its corollary under the assumption that the points given in the hypotheses are ruler points.

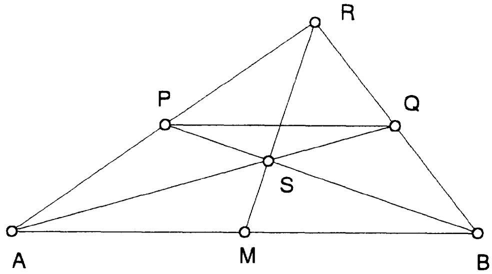


FIGURE 4.8.

Theorem 4.6. *The midpoint of two ruler points A and B is a ruler point if some other line parallel to $\overleftrightarrow{AB}$ is a ruler line. Conversely, if A,M,B,P are four ruler points with M the midpoint of A and B, then the line through P and parallel to $\overleftrightarrow{AB}$ is a ruler line.*

Figure 4.8 illustrates both constructions related to Theorem 4.6. Given a ruler line parallel to $\overleftrightarrow{AB}$, there are ruler points P and Q on that line such that $\square APQB$ is a trapezoid by Lemma 4.5. Then points R,S,M are determined in that order with M the desired midpoint of A and B. On the other hand, given ruler points A,M,B,P with M the midpoint of A and B and with P off $\overleftrightarrow{AB}$, then by Lemma 4.5 there is a ruler point R on $\overleftrightarrow{AP}$ that is distinct from A and P and such that $\overleftrightarrow{MR} \nparallel \overleftrightarrow{BP}$. Then S and Q are determined in that order, and $\overleftrightarrow{PQ}$ is the desired parallel. These two constructions are fundamental to ruler theory.

Theorem 4.7 (*cf.* Euclid I.10 and I.31). *The midpoint of two ruler points is a ruler point. The line through a ruler point and parallel to a ruler line is a ruler line.*

Proof. We prove the second part first. Since every ruler line contains three ruler points such that one is the midpoint of the other two by Lemma 4.5, then the lines through a ruler point and parallel to a ruler line are ruler lines by the second part of Theorem 4.6. The first part of Theorem 4.7 is now a consequence of the first part of Theorem 4.6. ∎

Since a ruler line contains two ruler points and since the midpoint of two ruler points is a ruler point, it follows that a ruler line contains infinitely many ruler points.

We next show that with the aid of Theorem 4.7 we can "add and subtract" segments on one of two given parallel lines. This is the first of two theorems that are informally known as the *Push-up-and-Pull-down Theorems*. In the First Push-up-and-Pull-down Theorem, segments on $\overleftrightarrow{PQ}$ are pushed up to congruent segments on a line parallel to $\overleftrightarrow{PQ}$ and then pulled down to congruent segments on $\overleftrightarrow{PQ}$ again.

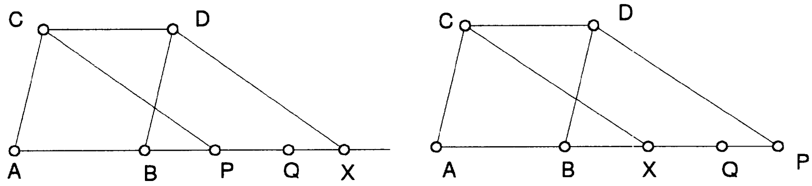


FIGURE 4.9.

Theorem 4.8. *Suppose P and Q are two ruler points. If A and B are two ruler points on $\overleftrightarrow{PQ}$, then the point X on $\overleftrightarrow{PQ}$ such that $PX = AB$ is a ruler point.*

Proof. Let C be a ruler point off $\overleftrightarrow{PQ}$. (At least one of the points in the starter set is off $\overleftrightarrow{PQ}$.) Let D be the ruler point that is the intersection of the ruler line through C that is parallel to $\overleftrightarrow{PQ}$ and the ruler line through B that is parallel to $\overleftrightarrow{AC}$. So $\square ACDB$ is a parallelogram. See Figure 4.9. If D and Q are on the same side of $\overleftrightarrow{CP}$, then let X be the intersection of $\overleftrightarrow{PQ}$ and the ruler line through D that is parallel to $\overleftrightarrow{CP}$. In this case, $\square PCDX$ is a parallelogram. If D and Q are on opposite sides of $\overleftrightarrow{CP}$, then let X be the intersection of $\overleftrightarrow{PQ}$ and the ruler line through C that is parallel to $\overleftrightarrow{DP}$. In this case, $\square XCDP$ is a parallelogram. In either case, X is a ruler point on $\overleftrightarrow{PQ}$ and $AB = CD = PX$, as desired. ∎

By pushing up segments on the X-axis to congruent segments on the line with equation $Y = 1$ and pulling them back down to congruent segments on the X-axis again, we have the following corollary.

Corollary 4.9. *If p and q are ruler numbers, then $p+q$ and $p-q$ are ruler numbers.*

The "multiplication and division" of segments will follow from the Second Push-up-and-Pull-down Theorem, which constructs a fourth proportional. In Figure 4.10, segments $\overline{OA}$ and $\overline{OB}$ are pushed up to proportional segments $\overline{DE}$ and $\overline{DG}$ through point F and then these segments are pulled down from point H to proportional segments $\overline{OC}$ and $\overline{OX}$.

Theorem 4.10 (*cf.* Euclid VI.12). *Suppose O and A are distinct ruler points. If ruler points B and C on $\overrightarrow{OA}$ are distinct from O, then the point X on $\overrightarrow{OA}$ such that $OA/OB = OC/OX$ is a ruler point.*

Proof. Let D be a ruler point off $\overleftrightarrow{OA}$. The line through D and parallel to $\overleftrightarrow{OA}$ is a ruler line. Since this line contains at least three ruler points different from D by Theorem 4.7, then there is a ruler point E on this line such that $\overleftrightarrow{AE} \nparallel \overleftrightarrow{OD}$ and $\overleftrightarrow{CE} \nparallel \overleftrightarrow{OD}$. Then, as in Figure 4.10, let X be defined by

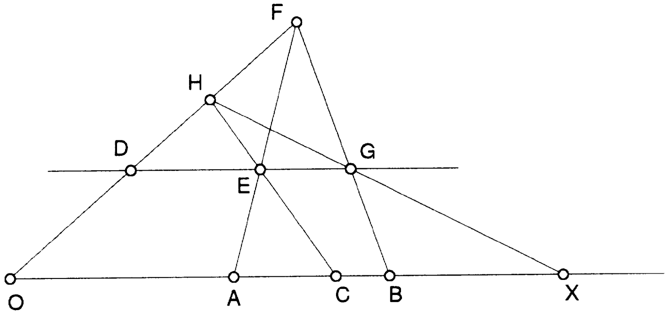


FIGURE 4.10.

$\overleftrightarrow{AE}, \overleftrightarrow{OD}$	$\overleftrightarrow{BF}, \overleftrightarrow{DE}$	$\overleftrightarrow{CE}, \overleftrightarrow{OD}$	$\overleftrightarrow{HG}, \overleftrightarrow{OA}$	X
F	G	H	X	

Therefore X is a ruler point. Recalling Figure 4.3b, we see that $OA/OB = DE/DG = OC/OX$, as desired. ∎

Corollary 4.11. *If p,q,r are ruler numbers and $r \neq 0$, then pq and p/r are ruler numbers.*

Proof. In the theorem we have $OX = (OB)(OC)/OA$. Suppose p,q,r are positive. Taking $O = (0,0)$, $A = (1,0)$, $B = (p,0)$, and $C = (q,0)$ in the theorem, we have $OX = pq$ and so pq is a ruler number; taking $O = (0,0)$, $A = (r,0)$, $B = (p,0)$, and $C = (1,0)$ in the theorem, we have $OX = p/r$ and so p/r is a ruler number. The corollary now follows since we know that $-x$ is a ruler number if x is a ruler number by Corollary 4.9. ∎

We come now to the principal theorem of ruler construction theory.

Theorem 4.12. *A point is a ruler point iff the coordinates of the point are rational numbers.*

Proof. Point $(1,0)$ is a ruler point. Hence $(n,0)$ is a ruler point for any integer n by Corollary 4.9. By Corollary 4.11, then $(m/n,0)$ is a ruler point for any integers m and n with $n \neq 0$. So $(r,0)$ is a ruler point for any rational number r. The Y-axis is a ruler line. That $(0,r)$, which is the intersection of the Y-axis and the line through $(r,0)$ that is parallel to the ruler line with equation $X + Y = 1$, is then a ruler point for any rational r follows from the part of Theorem 4.7 that corresponds to Euclid I.31. By the same theorem, the line with equation $X = r$, which passes through $(r,0)$ and is parallel to the Y-axis, and the line with equation $Y = s$, which passes through $(0,s)$ and is parallel to the X-axis, are ruler lines for any rational numbers r and s. If r and s are rational numbers, then the ruler lines with equations $X = r$ and $Y = s$ intersect at a ruler point. Therefore,

(r, s) is a ruler point for any rational numbers r and s. The converse is left for Exercise 4.6: If (r, s) is a ruler point, then both r and s must be rational numbers. ∎

Corollary 4.13. *The field of ruler numbers is* $\mathbb{Q}$.

In the proof of Theorem 4.12, we "rotated" the point $(r, 0)$ through an angle of 90° to the point $(0, r)$. Do you think we can rotate a segment through an angle of 45° with a ruler? Let $A = (0,0)$ and $B = (1,1)$. So $AB = \sqrt{2}$. If we could rotate $\overline{AB}$ clockwise 45° about$(0,0)$, then we could construct the point $(\sqrt{2}, 0)$ with the ruler. However, we know this is impossible since $\sqrt{2}$ is not a rational number. We can rotate $\overrightarrow{AB}$ through the 45° angle but not $\overline{AB}$. It is possible that AB is not a ruler number even though both A and B are ruler points.

By using the algebra of the rationals, we can now tell exactly what constructions are possible with the ruler alone. For one example, consider the analogues to Euclid I.11 and I.12, which ask for the construction of perpendiculars to a given line from a given point on the line and from a given point off the line. In Exercise 4.7, you are asked to find the equation of the perpendicular from ruler point (p, q) to the ruler line with equation $aX + bY + c = 0$. From this algebra, we find the perpendicular has an equation with rational coordinates. The next theorem then follows easily.

Theorem 4.14 (*cf.* Euclid I.11 and I.12). *The perpendicular through a ruler point to a ruler line is a ruler line.*

Of course the algebraic argument above does not tell you how to find a nice construction for such a perpendicular, merely that there is some construction. After spending a little time thinking about how to do this problem, you may appreciate the following construction. We suppose we have square $\square ABCD$ and want to construct the perpendicular from given point P to given line l. Let Q be the center of the square. Without loss of generality, we may suppose the line through Q that is parallel to l intersects $\overline{AB}$ at R. See Figure 4.11. Let S be the intersection of $\overline{CD}$ and the line through R that is parallel to $\overleftrightarrow{AD}$. So $AR = DS$. Let T be the intersection of $\overline{BC}$ and the line through S that is parallel to the diagonal $\overline{BD}$. So $DS = BT$. Hence, $AR = BT$. Considering the rotation about Q that sends A to B, we see that $\overline{QR}$ is rotated 90° onto $\overline{QT}$. In other words, $\overline{QT} \perp \overline{QR}$. Hence the line m through P that is parallel to $\overleftrightarrow{QT}$ is the desired perpendicular to l. Undoubtedly this construction is enjoyed more when you do not actually have to construct all those parallels with a ruler alone. In any case, you should enjoy reading through the construction.

Since our starter set allows us to construct parallels and perpendiculars with the ruler, we can also copy angles. This is another construction that is enjoyed best by only reading through it and not actually making a construction drawing. You will want to make your own sketch though,

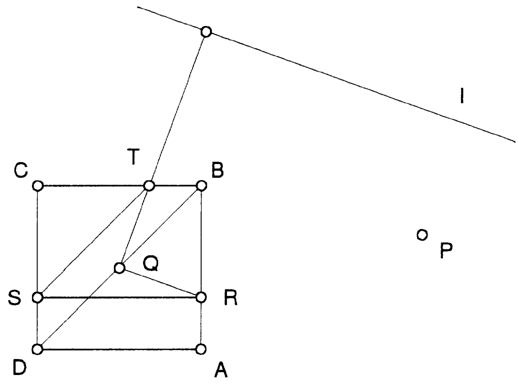


FIGURE 4.11.

drawing parallels and perpendiculars freehand. Given $\overrightarrow{VA}$ and $\angle PQR$, we want to construct point X such that $\angle AVX \cong \angle PQR$. We may suppose $\angle PQR$ is not a right angle as then the problem reduces to constructing a perpendicular. (See Exercise 4.13 for doubling an angle.) Construct B and C such that $\overleftrightarrow{VB} \parallel \overleftrightarrow{PQ}$ and $\overleftrightarrow{VC} \parallel \overleftrightarrow{QR}$. Then either $\angle BVC$ or a supplementary angle is congruent to $\angle PQR$. By using parallels, we have been able to reduce the problem to rotating $\angle BVC$ about its vertex. Since we can construct perpendiculars, we may suppose C is such that $\overleftrightarrow{BC} \perp \overleftrightarrow{VC}$ without loss of generality. If either one of $\overleftrightarrow{BV}$ or $\overleftrightarrow{VC}$ is perpendicular to $\overleftrightarrow{VA}$, then the perpendicular to the other at V will finish the problem. Hence, since vertical angles formed by intersecting lines are congruent, we may as well also suppose A is such that $\overleftrightarrow{BA} \perp \overleftrightarrow{VA}$. See Figure 4.12. Construct D, the foot of the perpendicular through V to $\overleftrightarrow{AC}$. Points V,A,B,C are concyclic. (Points A and C are on the circle with diameter $\overline{VB}$ by the converse of the Theorem of Thales.) So, $\angle VBC$ and $\angle VAC$ are congruent or supplementary. Since $\angle VAC$ and $\angle VAD$ are equal or supplementary, then $\angle BVC$ and $\angle AVD$ are congruent or supplementary. One of the cases $X = D$ or X such that D–V–X will finish the problem and complete a construction for Theorem 4.15.

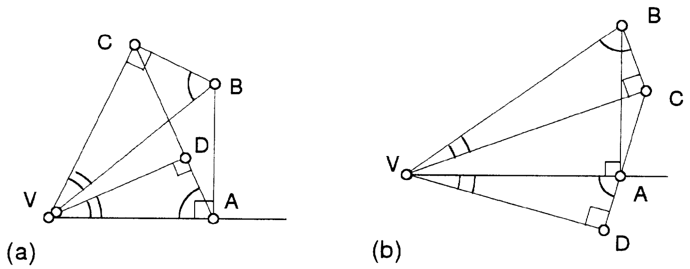


FIGURE 4.12.

Theorem 4.15 (*cf.* **Euclid I.23**). *If P,Q,R are noncollinear ruler points and if V and A are two ruler points, then there is a ruler point X such that $\angle AVX \cong \angle PQR$.*

With a ruler we can copy angles but, in general, we cannot copy segments. In the next chapter we shall have a tool that copies segments.

Exercises

4.1. Make construction drawings for the two constructions needed for Theorem 4.6.◊

4.2. Starting with only the seven points that are given in Figure 4.13 where $\square ABCD$ is a parallelogram, give a ruler construction for the line through P and parallel to l.◊

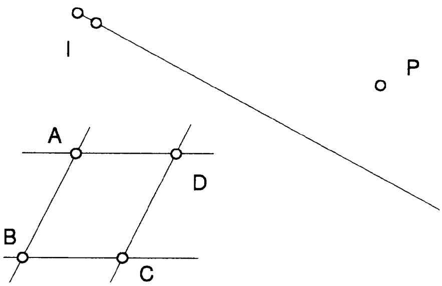


FIGURE 4.13.

4.3. Given the vertices of trapezoid $\square ABCD$ with $\overleftrightarrow{AB} \parallel \overleftrightarrow{CD}$, make a construction drawing for a ruler construction of point E such that A is the midpoint of E and B and of point F such that B is the midpoint of E and F.

4.4. Make construction drawings to illustrate Corollary 4.9.

4.5. Make construction drawings to illustrate Corollary 4.11.

4.6. Prove the "only if" part of Theorem 4.12: The coordinates of a ruler point are rational.◊

4.7. Find an equation of the perpendicular from ruler point (p, q) to the ruler line with equation $aX + bY + c = 0$.◊

4.8. State the steps in a ruler construction corresponding to Euclid I.46: Construct a square with a given segment as one side.◊

4.9. Show that every right angle that can be constructed with a ruler can be bisected with a ruler but not every angle that can be constructed with a ruler can be bisected with a ruler.◊

4.10. Suppose $0 < x < 90$. Prove that some angle of $x°$ can be constructed with a ruler iff $\tan x°$ is in $\mathbb{Q}$.◊

4.11. Show that it is impossible to construct an equilateral triangle with a ruler.◊

4.12. Suppose A and B are ruler points. Show that the point P on $\overrightarrow{AB}$ such that $AP = 1$ is not necessarily a ruler point.◊

4.13. Show the image of a ruler point under the reflection in a ruler line is a ruler point. Show that a given acute angle can be doubled with the ruler alone.◊

4.14. Prove Varignon's Theorem.◊

4.15. Give another starter set for Definition 4.3 that still produces Theorem 4.12.◊

4.16. If the vertices of $\triangle ABC$ are ruler points, then show that the centroid, the circumcenter, and the orthocenter of the triangle are ruler points. (The *orthocenter* of a triangle is the point of concurrence of the three altitudes of the triangle.)◊

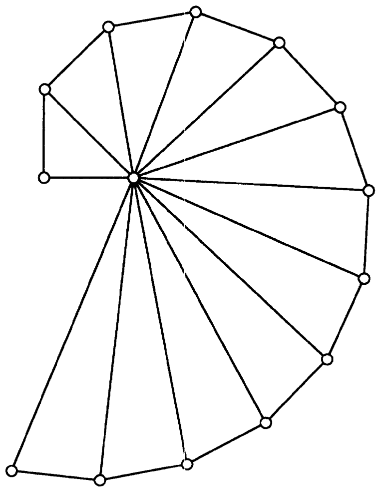
4.17. Show that the ruler line with equation $Y = X$ and the ruler circle with equation $X^2 + Y^2 = 1$ do not intersect at ruler points. Show that there are no ruler points on the circle with equation $X^2 + Y^2 = 3$.◊

4.18. Suppose that A and B are ruler points and that $\tan x°$ is rational with $0 < r < 90$. Is there a ruler point C such that $m\angle CAB = x$?◊

4.19. Give a ruler construction for the trisection of $\overline{OP}$, where $O = (0, 0)$ and $P = (a, 0)$ with a in $\mathbb{Q}$.◊

4.20. Suppose the five ruler points A,B,C,D,E are on one circle. Suppose F is a ruler point. Give a ruler construction for the intersection of $\overleftrightarrow{EF}$ and the circle.◊

5
The Ruler and Dividers



A ruler is used to draw lines. A dividers is used to carry distances. Both are everyday tools used by a drafter. If you have seen what looked like a compass with two points and no place for a pencil, that was a dividers. As a compass is sometimes called a pair of compasses, a dividers is often called a pair of dividers. Given points A,B,C,D with $A \neq B$, the dividers is used to construct the point P on $\overrightarrow{AB}$ such that $AP = CD$. Therefore, with the dividers we can copy segments. Can we swing the dividers with one end on a point until the other end comes to rest on a given line? No, only because it is against the rules here; we call the tool that does that a compass. Knowing the allowed use of the dividers, you should expect the

G. E. Martin, *Geometric Constructions*

definition below. For practice in formulating these definitions, you should write down the definition of a ruler and dividers point before you read any further.

Definition 5.1. In the cartesian plane, a point is a ***ruler and dividers point*** if the point is the last of a finite sequence P_1,P_2, ... ,P_n of points such that each point is in $\{(1,0), (0,1), (2,0), (0,2)\}$ or is obtained in one of two ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; and (ii) as the point P on $\overrightarrow{AB}$ such that $AP = CD$ where A,B,C,D are points that appear earlier in the sequence with $A \neq B$. A ***ruler and dividers line*** is a line that passes through two ruler and dividers points. A ***ruler and dividers circle*** is a circle through a ruler and dividers point with a ruler and dividers point as center. A number x is a ***ruler and dividers number*** if $(x, 0)$ is a ruler and dividers point.

There is no question that all ruler points are ruler and dividers points. There is a question about the ruler constructions being valid when a larger set of points is used. Technically, we are required to prove new analogues to Euclid I.10, I.11, I.12, and I.31, as well as the Push-up-and-Pull-down Theorems and whatever else we may want from the preceding chapter, with the larger set of ruler and dividers points in place of the ruler points in the statements of the theorems. We are not going to do that, however. The statements of the theorems are changed, but the proofs are essentially unchanged. This might be a good time to review the ruler constructions to see that they do carry over. For example, the same constructions that show the set of ruler numbers forms a field will show the larger set of ruler and dividers numbers also forms a field. We shall simply suppose all the theorems of the preceding chapter with the necessary changes required to modify them so that they apply here. We suppose you could provide a proof for any of the statements collected in Theorem 5.2.

Theorem 5.2. *The intersection of two ruler and dividers lines is a ruler and dividers point. If A,B,C,D are ruler and dividers points with $A \neq B$, then the point P on $\overrightarrow{AB}$ such that $AP = CD$ is a ruler and dividers point. The midpoint of two ruler and dividers points is a ruler and dividers point. The line through a ruler and dividers point and parallel to a ruler and dividers line is a ruler and dividers line. The perpendicular through a ruler and dividers point to a ruler and dividers line is a ruler and dividers line. The ruler and dividers numbers form a field.*

You can imagine what a rusty dividers is. Given two points A and B, the *rusty dividers* is used to construct the point P on $\overrightarrow{AB}$ such that $AP = 1$. The rusty dividers also has the less colorful names *gauge* and *scale*.

Another construction tool is the angle bisector. Needless to say, the idea behind the *angle bisector* is to bisect a given angle. This, by the way, is a tool that is easily made. The angle bisector, illustrated in Figure 5.1, is just

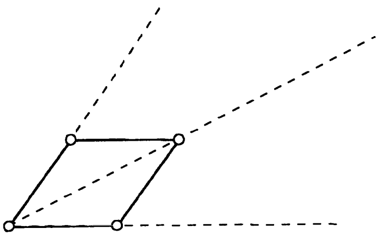


FIGURE 5.1.

a square with pivots at the vertices. When the angle bisector is inserted in an angle, one of its diagonals determines the desired angle bisector; the principle is that the diagonals of a rhombus bisect the angles of the rhombus. So, given $\angle AVB$, this angle bisector tool constructs the point P such that $\square PQVR$ is a unit rhombus with Q on $\overrightarrow{VA}$ and R on $\overrightarrow{VB}$.

The last of the construction tools to be introduced here is, believe it or not, the *cannon*. See Figure 5.2. This cannon cannot be moved, except to rotate about its center, and always shoots the cannon ball the same distance. The cannon sits at the origin and is aimed along a given ray from the origin to determine the point on the ray that is a unit distance from the origin. The cannon is a rusty dividers with one of its points always restricted to the origin. Perhaps the rusty dividers should be called a portable cannon. In any case, with $O = (0,0)$ and another given point A, the cannon is used to construct the point P on $\overrightarrow{OA}$ such that $OP = 1$.

Ruler and dividers constructions are among those discussed in Hilbert's *Foundations of Geometry.* This book has had a strong influence on twentieth-century mathematics, and David Hilbert (1862–1943) is generally considered the greatest mathematician of the twentieth century. In the history of geometry, only *Elements* by Euclid and *Geometry* by Descartes have been more influential than *Foundations of Geometry* by Hilbert.

By now you should be very familiar with the format of our definitions; if not, you will be after doing Exercise 5.1.

Definition 5.3. The definitions of a ***ruler and rusty dividers point***, a ***ruler and rusty dividers line***, a ***ruler and cannon point***, a ***ruler and cannon line***, a ***ruler and angle bisector point***, and a ***ruler and angle bisector line*** are left for Exercise 5.1.

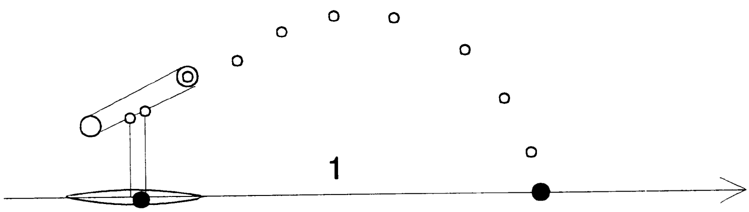


FIGURE 5.2.

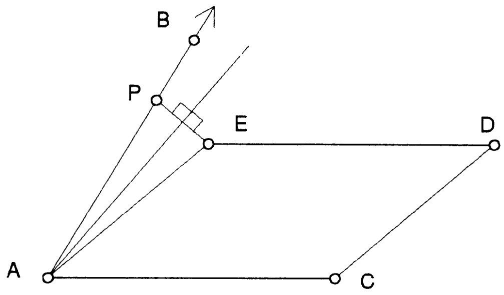


FIGURE 5.3.

If someone tells you how a wheelbarrow could be used in geometric constructions, you should know how to define a *ruler and wheelbarrow point* by the time you have finished Exercise 5.1, which you should do before reading the following theorem. From Theorem 5.4, it is clear why we are studying all these tools at once.

Theorem 5.4. *The following are equivalent:*

1. *Point P is a ruler and dividers point.*
2. *Point P is a ruler and angle bisector point.*
3. *Point P is a ruler and cannon point.*
4. *Point P is a ruler and rusty dividers point.*

Proof. (**1** $\Rightarrow$ **2**) We want to show that any point constructed with ruler and dividers can be constructed with ruler and angle bisector. We need to show that given points A,B,C,D with $A \neq B$ then point P on $\overrightarrow{AB}$ such that $AP = CD$ can be constructed with ruler and angle bisector. We may suppose $C \neq D$, as otherwise $P = A$. If $\overleftrightarrow{AB} \parallel \overleftrightarrow{CD}$, then P is constructed with the ruler alone by drawing parallelograms as in the First Push-up-and-Pull-down Theorem. If $\overleftrightarrow{AB} \perp \overleftrightarrow{CD}$, the perpendiculars through C and D to the angle bisector of one of the right angles formed by $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ intersect $\overleftrightarrow{AB}$ in two points C' and D' such that $C'D' = CD$ and we are back to the previous case. (So far the angle bisector has not been necessary since we can bisect a right angle with a ruler, although we cannot bisect a general angle.) We suppose now that $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ are neither parallel nor perpendicular. If $A = C$, let $E = D$. If $A \neq C$, point E is constructed with the ruler alone so that $\square ACDE$ is a parallelogram with E off $\overleftrightarrow{AB}$. See Figure 5.3. By our hypothesis, the angle bisector of $\angle EAB$ can be constructed. Then the perpendicular from E to the angle bisector is constructible with the ruler alone. The intersection of $\overrightarrow{AB}$ and the perpendicular to the angle bisector is the desired point P, since $AP = AE = CD$.

(**2** $\Rightarrow$ **3**) This time we must construct a particular point on the angle bisector of given angle $\angle AVB$ by using the ruler and cannon. See Figure 5.4.

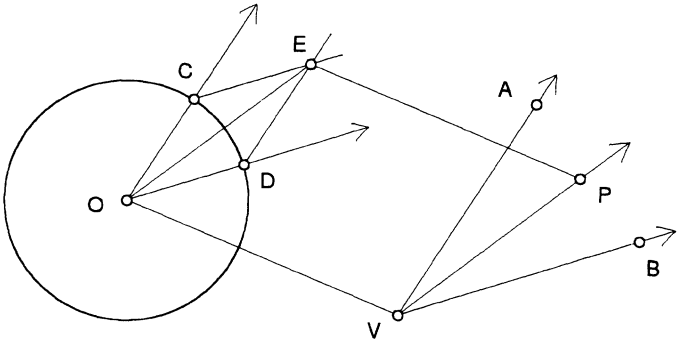


FIGURE 5.4.

With $O = (0,0)$, let C and D be points on the unit circle such that $\overrightarrow{OC} \parallel \overrightarrow{VA}$, $\overrightarrow{OD} \parallel \overrightarrow{VB}$, and $\angle COD \cong \angle AVB$. Lines $\overleftrightarrow{OC}$ and $\overleftrightarrow{OD}$ are constructible with the ruler alone. Then the points C and D are determined by the cannon. The point E such that $\square ECOD$ is a parallelogram is constructible with the ruler. So $\overrightarrow{OE}$ bisects $\angle COD$ and $\square ECOD$ is a unit rhombus. If V,O,E are not collinear, then point P such that $\square VOEP$ is a parallelogram is also constructible with the ruler alone. So $\overrightarrow{VP}$ bisects $\angle AVB$, and P is such that there exists a point Q on $\overrightarrow{VA}$ and a point R on $\overrightarrow{VB}$ such that $\square PQVR$ is a unit rhombus. (Note that we do not need to construct the points Q and R, although we could if we wanted to. This part of the proof may have to be altered to conform to your definition of a ruler and angle bisector point as given in Exercise 5.1.) Finally, if V,O,E are collinear, then the desired point P can be obtained with the ruler alone by the First Push-up-and-Pull-down Theorem.

(**3** $\Rightarrow$ **4**) and (**4** $\Rightarrow$ **1**) are trivial. ∎

We introduce a symbol for the field of ruler and dividers numbers, which are sometimes called *pythagorean numbers* for a reason that will soon become apparent.

Definition 5.5. The field of ruler and dividers numbers is denoted by $\mathbb{P}$.

Theorem 5.6. *If a,b,c are in $\mathbb{P}$ with $c \neq 0$, then*

$$a+b, \quad a-b, \quad ab, \quad a/c, \quad \sqrt{a^2+b^2}$$

are in $\mathbb{P}$. If P and Q are ruler and dividers points, then PQ is a ruler and dividers number.

Proof. The four arithmetic terms come from the ruler theory; we have already stated that $\mathbb{P}$ is a field. It is the presence of the radical that requires proof. If a and b are ruler and dividers numbers, then (a,b) is a ruler and dividers point P. Let $O = (0,0)$ and $U = (1,0)$. Then there is a ruler and dividers point X on $\overrightarrow{OU}$ such that $OX = OP$. Since $OP = \sqrt{a^2+b^2}$, then $X = (\sqrt{a^2+b^2}, 0)$ and the radical is in $\mathbb{P}$. The second statement in the theorem now follows from the familiar formula $\sqrt{(c-a)^2+(d-b)^2}$ for the distance between the two points (a, b) and (c, d) in the cartesian plane. ■

We know that $\mathbb{Q} \subset \mathbb{P} \subseteq \mathbb{E} \subset \mathbb{R}$. Every ruler and dividers number is a ruler and compass number. To show the converse, we would need to show that every ruler and compass point is a ruler and dividers point. The converse, however, turns out to be false. We shall see that $\mathbb{E}$ is larger than $\mathbb{P}$. A number in $\mathbb{P}$ is also in $\mathbb{E}$ and, therefore, is in an iterated quadratic extension of $\mathbb{Q}$. However, each of these quadratic extensions is of a special kind. Let's take a closer look.

It is algebraically easier to look at the ruler and dividers points as ruler and cannon points. So, suppose P is a ruler and cannon point. We will mimic the proof of Theorem 2.15, which shows that a ruler and compass point has coordinates in an iterated quadratic extension of $\mathbb{Q}$. From the definition of a ruler and cannon point, we see that P must be the last of a sequence $P_1, P_2, \ldots, P_n$ of points, each of which is in $\{(1,0), (0,1), (2,0), (0,2)\}$ or is obtained in one of two ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; and (ii) as the point P on $\overrightarrow{OA}$ such that $OP = 1$ where $O = (0,0)$ and A is a point that appears earlier in the sequence with $A \neq O$. By the sequence of lemmas preceding Theorem 2.15, we can associate P_1 with the rationals and observe that each point P_i for $i > 1$ can be associated with a field F_i such that the coordinates of P_i are in F_i and such that F_i is either equal to F_{i-1} or else is a quadratic extension of F_{i-1}. If P_i is obtained by (i) above, then $F_i = F_{i-1}$. Suppose P_i is determined by (ii) above with $A = (a, b)$. Then P_i is $(a/\sqrt{a^2+b^2}, b/\sqrt{a^2+b^2})$. If a^2+b^2 is a square in F_{i-1}, then the coordinates of P_i are in F_{i-1} and $F_i = F_{i-1}$. However, if a^2+b^2 is not a square in F_{i-1}, then the coordinates of P_i are in a quadratic extension of F_{i-1} and $F_i = F_{i-1}(\sqrt{a^2+b^2})$. We conclude that every ruler and dividers number is in an iterated quadratic extension of $\mathbb{Q}$, where all the quadratic extensions are of a special type. Before stating our conclusion, we need a definition of this special type of quadratic extension.

Definition 5.7. If x and y are elements of field F but x^2+y^2 is not a square in F, then $F(\sqrt{x^2+y^2})$ is a ***pythagorean extension of*** F. If F_{i+1} is a pythagorean extension of F_i for $i = 1, 2, \ldots, n-1$, then each of $F_1, F_2, \ldots, F_n$ is an ***iterated pythagorean extension of*** F_1. A field is ***pythagorean*** if every sum of squares in the field is a square in the field.

By the argument preceding the definition, every ruler and dividers number is in an iterated pythagorean extension of the rationals. By Theorem 5.6, every number in an iterated pythagorean extension of $\mathbb{Q}$ is in $\mathbb{P}$.

Therefore, we have already proved the following important theorem, which characterizes the coordinates of the ruler and dividers points.

Theorem 5.8. *A real number is a ruler and dividers number iff the number is an element of an iterated pythagorean extension of* $\mathbb{Q}$.

There are quadratic extensions that are not pythagorean extensions. We can check the answer to Exercise 4.17 in The Back of the Book to find an argument that shows there are no rational numbers x and y such that $x^2 + y^2 = 3$. Hence, $\mathbb{Q}(\sqrt{3}\,)$ is a quadratic extension of $\mathbb{Q}$ that is not a pythagorean extension of $\mathbb{Q}$. This is not to say that $\sqrt{3}$ is not in $\mathbb{P}$. Since $2 = 1^2 + 1^2$, then $\mathbb{Q}(\sqrt{2}\,)$ is a pythagorean extension of $\mathbb{Q}$. Since $3 = (\sqrt{2}\,)^2 + 1^2$, then $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$ is a pythagorean extension of $\mathbb{Q}(\sqrt{2}\,)$. So $\sqrt{3}$ is in the iterated pythagorean extension $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)$ of $\mathbb{Q}$ and therefore in $\mathbb{P}$.

As $\mathbb{E}$ is the union of all iterated quadratic extensions of $\mathbb{Q}$, so $\mathbb{P}$ is the union of all iterated pythagorean extensions of $\mathbb{Q}$. Note that $\sqrt{a^2+b^2} = \pm a\sqrt{(1+(b/a)^2}$ if $a \neq 0$. Thus, as $\mathbb{E}$ is obtained from the rationals by repeated application of the five operations $+$, $-$, $\times$, $\div$, and $\sqrt{\ }$, so $\mathbb{P}$ is obtained from the rationals by repeated application of the five operations $+$, $-$, $\times$, $\div$, and $\sqrt{(1+(\ )^2)}$. Field $\mathbb{E}$ is pythagorean since a sum of squares is positive or zero. It is certainly not obvious that $\mathbb{P}$ and $\mathbb{E}$ are distinct.

Theorem 5.9. *Field* $\mathbb{P}$ *is a pythagorean field.*

Proof. By Theorem 5.6, field $\mathbb{P}$ is such that if x and y are in $\mathbb{P}$, then $\sqrt{(x^2+y^2)}$ is in $\mathbb{P}$. Thus we need show only that "every sum of squares" in the definition of a pythagorean field above is equivalent to "every sum of two squares." This follows from the identities

$$\sqrt{c_1^2+c_2^2+c_3^2} = \sqrt{\sqrt{c_1^2+c_2^2}^{\,2} + c_3^2},$$

$$\sqrt{c_1^2+c_2^2+c_3^2+c_4^2} = \sqrt{\sqrt{c_1^2+c_2^2+c_3^2}^{\,2} + c_4^2},$$

$$\sqrt{c_1^2+c_2^2+c_3^2+c_4^2+c_5^2} = \sqrt{\sqrt{c_1^2+c_2^2+c_3^2+c_4^2}^{\,2} + c_5^2},$$

$$\vdots$$

$$\sqrt{c_1^2+c_2^2+\ldots+c_{n-1}^2+c_n^2} = \sqrt{\sqrt{c_1^2+c_2^2+\ldots+c_{n-1}^2}^{\,2} + c_n^2}.$$

Hence $\mathbb{P}$ is pythagorean. ∎

As $\mathbb{E}$ is the smallest euclidean field, so $\mathbb{P}$ is the smallest pythagorean field. We now set out to prove that $\mathbb{P}$ and $\mathbb{E}$ are distinct.

In case there is a question, a square in a field F is considered to be a sum of squares in F; we can write $a^2 = a^2 + 0^2$. Not only are sums and products of sums of squares in a field F sums of squares in F, but, perhaps surprisingly, a quotient of sums of squares in a field F is a sum of squares in F. The following examples from $\mathbb{Q}$ precede the general theorem:

$$\begin{aligned}(2^2+3^2)+(5^2+7^2) &= 2^2+3^2+5^2+7^2,\\ (2^2+3^2)\times(5^2+7^2) &= 2^25^2+2^27^2+3^25^2+3^27^2\\ &= 10^2+14^2+15^2+21^2,\\ (2^2+3^2)\div(5^2+7^2) &= [(2^2+3^2)(5^2+7^2)][1/(5^2+7^2)^2]\\ &= (10/74)^2+(14/74)^2+(15/74)^2+(21/74)^2.\end{aligned}$$

If a and b are positive integers, then the rational a/b can always be written as the sum in $\mathbb{Q}$ of ab squares, each of which is $(1/b)^2$.

Theorem 5.10. *If x,y,z are sums of squares in field F and $z \neq 0$, then*

$$x+y, \quad xy, \quad \textit{and} \quad x/z$$

are sums of squares in F.

Proof. The result is clear for sums and products. Since $x/z = [xz][(1/z)^2]$, then the quotient x/z of sums of squares in F is a product of two sums of squares in F and so a sum of squares in F. ∎

Every nonnegative number in $\mathbb{Q}$ is a sum of squares. This is generally not the case for a field. Of course, none of the numbers $5-16\sqrt{2}$, $-5-16\sqrt{2}$, $3-40\sqrt{2}$, and $-\sqrt{2}$ can be a sum of squares in $\mathbb{Q}(\sqrt{2})$ because no negative number can be a sum of squares in $\mathbb{R}$. The next theorem will prove that therefore none of the corresponding positive numbers $5+16\sqrt{2}$, $-5+16\sqrt{2}$, $3+40\sqrt{2}$, and $\sqrt{2}$ can be a sum of squares in $\mathbb{Q}(\sqrt{2})$.

Theorem 5.11. *Suppose x,y,d are in field F, $d>0$, but $\sqrt{d}$ is not in F. Then $x+y\sqrt{d}$ is a sum of squares in $F(\sqrt{d})$ iff $x-y\sqrt{d}$ is a sum of squares in $F(\sqrt{d})$.*

Proof. Suppose $x+y\sqrt{d} = \sum (x_i + y_i\sqrt{d})^2$. So

$$x+y\sqrt{d} = \sum (x_i^2 + y_i^2 d + 2x_iy_i\sqrt{d}) = [\sum (x_i^2+y_i^2d)] + [2\sum x_iy_i]\sqrt{d}.$$

Then $y = 2\sum x_iy_i$, since $\sqrt{d}$ is not in F, and so $x = \sum(x_i^2+y_i^2d)$. Hence, $x-y\sqrt{d} = \sum(x_i - y_i\sqrt{d})^2$, as desired. The converse follows by changing the sign of y. ∎

Corollary 5.12. *Suppose x,y,d are in field F, $x \leq 0$, $y \neq 0$, $d>0$, but $\sqrt{d}$ is not in F. Then $x+y\sqrt{d}$ is not a sum of squares in $F(\sqrt{d})$.*

Proof. Since one of the numbers $x+y\sqrt{d}$ and $x-y\sqrt{d}$ is negative and therefore cannot be a sum of squares even in $\mathbb{R}$, then neither of the numbers is a sum of squares in $F(\sqrt{d})$. ∎

The corollary can also be proved without reference to the theorem, since $\sum (x_i^2 + y_i^2 d)$ cannot be negative. The next theorem is the key proposition in showing that $\mathbb{P} \neq \mathbb{E}$.

Theorem 5.13. *If G is a pythagorean extension of field F and number z in F is a sum of squares in G, then z is a sum of squares in F.*

Proof. Let $G = F(\sqrt{d})$. Suppose z is in F and is a sum of squares in G. Then, since $\sqrt{d}$ is not in F, there are x_i and y_i in F such that

$$\begin{aligned} z = z + 0\sqrt{d} = \sum (x_i + y_i\sqrt{d})^2 &= [\sum (x_i^2 + y_i^2 d)] + [2\sum x_i y_i]\sqrt{d} \\ &= \sum (x_i^2 + y_i^2 d). \end{aligned}$$

Because d is a sum of squares in F, then $\sum (x_i^2 + y_i^2 d)$ is a sum of squares in F by Theorem 5.10. So z is a sum of squares in F. ∎

Theorem 5.14. *If G is an iterated pythagorean extension of field F and number z in F is a sum of squares in G, then z is a sum of squares in F.*

Proof. The theorem follows by repeated application of the preceding theorem, starting at the top of a tower of pythagorean extensions and working down to F. ∎

This theorem can be recast into the following language. A pythagorean number "lives" in $\mathbb{P}$. If you are a pythagorean number that lives in a tower of iterated pythagorean extensions of field F, then you were either born a sum of squares (*i.e.*, you were a sum of squares on the floor on which you first appeared) or else you have no chance of becoming a sum of squares later in life on some higher floor of the tower. In this analogy, a number lives on the floor it was born and every floor above in this tower. Although a pythagorean number is born and lives in many towers, any two can be combined, in the sense of Exercise 5.15, into one tower. Therefore, wherever a pythagorean number is born it is a sum of squares at birth or else is not a sum of squares in any other tower or even in $\mathbb{P}$. The Pythagorean numbers seem to live in a very aristocratic world with respect to becoming a sum of squares.

Theorem 5.15. $\mathbb{P} \neq \mathbb{E}$.

Proof. Number $1 + \sqrt{5}$ is in both $\mathbb{P}$ and $\mathbb{E}$ since $5 = 1^2 + 2^2$. Clearly, $\sqrt{1 + \sqrt{5}}$ is in $\mathbb{E}$. Assume $\sqrt{1 + \sqrt{5}}$ is also in $\mathbb{P}$. Then $1 + \sqrt{5}$ is a sum of two squares in some iterated pythagorean extension $\mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_n})$ of $\mathbb{Q}$. Let $G_0 = \mathbb{Q}(\sqrt{5})$. We define the field G_i for $i = 1, 2, \ldots, n$ as follows. Let G_i be the pythagorean extension $G_{i-1}(\sqrt{d_i})$ of G_{i-1} unless $\sqrt{d_i}$ is already in G_{i-1}, in which case let $G_i = G_{i-1}$. Since $\mathbb{Q}(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_i})$ is a subset of G_i for $i = 1, 2, \ldots, n$, then $1 + \sqrt{5}$ is a sum of squares in the iterated pythagorean extension G_n of G_0. By Theorem 5.14, then

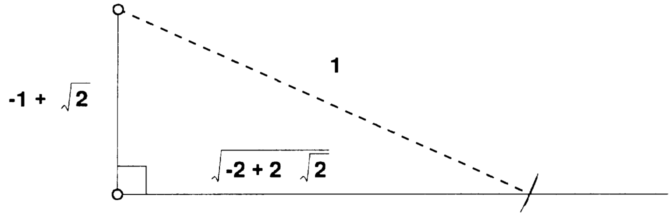


FIGURE 5.5.

$1 + \sqrt{5}$ is a sum of squares in G_0. Since $G_0 = \mathbb{Q}(\sqrt{5}\,)$, we now have the contradiction that the negative number $1-\sqrt{5}$ is a sum of squares in $\mathbb{Q}(\sqrt{5}\,)$, by Theorem 5.11. Therefore, $\sqrt{1+\sqrt{5}}$ is in $\mathbb{E}$ but not in $\mathbb{P}$. ∎

In the proof above we could have used $\sqrt{2}$ in place of $1+\sqrt{5}$ and argued that $\sqrt[4]{2}$ is in $\mathbb{E}$ but not in $\mathbb{P}$. After you have studied the proof, Corollary 5.12 should easily provide many possible positive numbers that are not sums of squares at birth and whose square roots therefore live in $\mathbb{E}$ but not in $\mathbb{P}$. For example, $\sqrt{-2+2\sqrt{2}}$ is in $\mathbb{E}$ but not in $\mathbb{P}$.

Although segments of length 1 and $-1+\sqrt{2}$ can be constructed by ruler and dividers, it takes the power of a compass to construct a segment of length $\sqrt{-2+2\sqrt{2}}$. See Figure 5.5. From this same figure, we can see that knowing how to construct a segment the length of the cosine of an angle is not sufficient for constructing the angle with ruler and dividers. Constructing the tangents to a circle from a point outside the circle is generally impossible with ruler and dividers. For example, if $A = (0,0)$, $B = (-1+\sqrt{2}, 0)$, and $P = (1,0)$, then the tangent to A_B that passes through P cannot be constructed with ruler and dividers.

We turn next to a famous ruler and compass construction problem. As we will see, the problem can be solved with ruler and dividers. See Figure 5.6.

Malfatti's Problem. Inscribe in a given triangle the three circles such that each circle is externally tangent to the other two and each circle is tangent to two sides of the triangle.

In 1803, Professor Gianfrancesco Malfatti (1731–1807) of the University of Ferrara posed the following question: Given a triangular prism of any sort of material, such as marble, how shall three circular cylinders of the same height as the prism and of the greatest possible volume be related to one another in the prism and leave over the least possible amount of material? This question reduces to the plane construction problem that we will refer to here as the *marble problem*: How do you cut three circles from a given triangle so that the sum of the areas of the three circles is maximized? Malfatti was only the first to assume that what has come to be called Malfatti's Problem and the marble problem have the same solution for a given

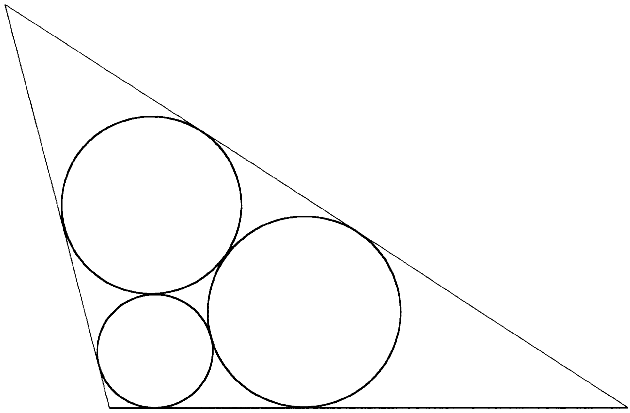


FIGURE 5.6.

triangle. In spite of the popularity of Malfatti's Problem, it was not until 1929, which was over 125 years after the problem was first published, that Lob and Richmond pointed out that for an equilateral triangle the solution to the marble problem is suggested by Figure 5.7 and not Figure 5.6. Even more surprising, the American geometer Michael Goldberg showed in 1967 that for a given triangle the Malfatti's Problem and the marble problem never have the same solution. In each of the triangles in Figure 5.7, the first circle is taken as the inscribed circle of the given triangle; a second circle is then inscribed in the smallest angle of the triangle and externally tangent to the first circle; and a third circle is then appropriately inscribed in the same angle or the next larger angle of the triangle, whichever allows the larger circle. Goldberg showed that for any triangle this construction gives three circles with larger total area than that given by the three circles that solve Malfatti's Problem for the same triangle. Analytic conditions for the solution to the marble problem can be found in the 1994 paper by Zalgaller and Los′. Constructions for Figure 5.7 are easy, but a construction for Figure 5.6 is far from trivial.

In 1826, Professor Jacob Steiner (1796–1863) of the University of Berlin published a singularly simple solution to Malfatti's Problem. The construction by this great geometer, who was born the son of a Swiss farmer, stimu-

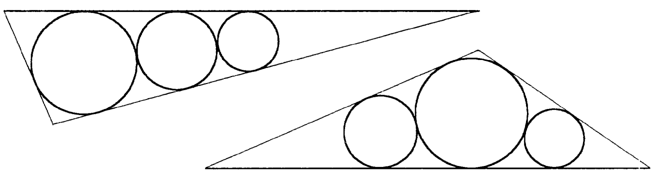


FIGURE 5.7.

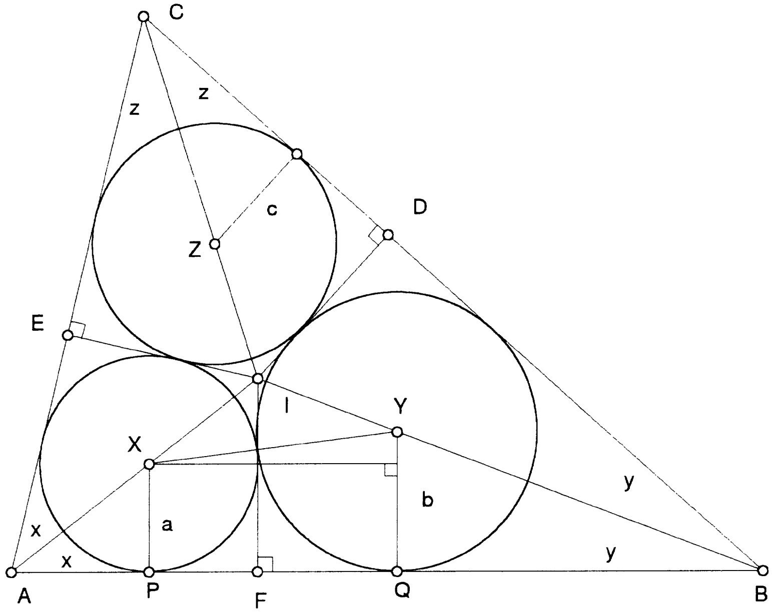


FIGURE 5.8.

lated interest in the problem because his construction was published without a proof. Such is also the case here. See Exercise 5.20.

In solving Malfatti's Problem, we will use over and over again the fact that the tangents from a point outside a circle to the circle are congruent. The notation is given in Figure 5.8. Points D,E,F are the feet of the perpendiculars to the indicated sides from I, the incenter of given $\triangle ABC$. Then $r = ID = IE = IF$, where r is the radius of the inscribed circle. Points X,Y,Z are the centers of the desired circles, which have radii a,b,c. Let $s = AE = AF$, $t = BF = BD$, and $u = CD = CE$. Let x,y,z be the lengths of the tangents to the adjacent circles from the vertices of $\triangle ABC$. Let P and Q be on $\overline{AB}$ such that $x = AP$ and $y = BQ$.

Since $PQ = \sqrt{(a+b)^2 - (a-b)^2} = 2\sqrt{ab}$, then we can rewrite the equation $AB = AP + PQ + QB$ as $s + t = x + 2\sqrt{ab} + y$. Now, $a/x = r/s$ since $\triangle XPA \sim \triangle IFA$ and, likewise, $b/y = r/t$. So $ab = r^2(xy/st)$. Then, we rewrite the equation above again, this time as the first of the following three equations:

$$x + y + 2r\sqrt{xy/st} = s + t,$$
$$y + z + 2r\sqrt{yz/tu} = t + u,$$
$$z + x + 2r\sqrt{zx/us} = u + s.$$

The second and third of these equations follow by symmetry. We note that r,s,t,u are ruler and dividers numbers determined by $\triangle ABC$ and that the problem is solved once x,y,z are found. Fortunately, Malfatti solved these three equations for us and got

$$2x = s + t + u - r + \sqrt{r^2 + s^2} - \sqrt{r^2 + t^2} - \sqrt{r^2 + u^2},$$
$$2y = s + t + u - r - \sqrt{r^2 + s^2} + \sqrt{r^2 + t^2} - \sqrt{r^2 + u^2},$$
$$2z = s + t + u - r - \sqrt{r^2 + s^2} - \sqrt{r^2 + t^2} + \sqrt{r^2 + u^2}.$$

All we have to do is convince ourselves that there is a unique solution to the problem and check that Malfatti's solution does work. We also have to note the somewhat surprising property that x,y,z are ruler and dividers numbers. Thus, Malfatti's Problem is solvable with ruler and dividers! A little more algebra yields another form of the solution. Since $IA = \sqrt{r^2 + s^2}$, $IB = \sqrt{r^2 + t^2}$, $IC = \sqrt{r^2 + u^2}$, and $s + t + u$ is half the perimeter of $\triangle ABC$, we let $2m = IA + IB + IC + r - (AB + BC + CA)/2$ and get

$$x = IA - m,$$
$$y = IB - m,$$
$$z = IC - m.$$

Exercises

5.1. Give all the definitions required for Definition 5.3.◊

5.2. Explain in detail why a euclidean field is a pythagorean field.◊

5.3. Give a ruler and dividers construction corresponding to Euclid I.31.◊

5.4. Give a ruler and rusty dividers construction corresponding to Euclid I.9, which asks for the angle bisector of a given angle.◊

5.5. With the hint that the altitudes of a triangle are concurrent, give a ruler and rusty dividers construction for a perpendicular to given line l.◊

5.6. Show $\mathbb{Q}(\sqrt{7})$ is a subset of $\mathbb{P}$; show $\mathbb{Q}(\sqrt{7/2})$ is a subset of $\mathbb{P}$.◊

5.7. Show that any quadratic extension of $\mathbb{Q}$ is a subset of $\mathbb{P}$.◊

5.8. Suppose point C is given off $\overleftrightarrow{AB}$. Sketch a ruler and dividers construction for the quinquesection of $\overline{AB}$. Sketch a ruler and rusty dividers construction for the trisection of $\overline{AB}$.◊

5.9. With A,B,C,D ruler and dividers points, show that the points in the intersection of A_B and $\overleftrightarrow{AD}$ are ruler and dividers points, that the points in the intersection of A_B and $\overleftrightarrow{BD}$ are ruler and dividers points, but that the points in the intersection of A_B and $\overleftrightarrow{CD}$ may fail to be ruler and dividers points.◊

5.10. Suppose $0 < x < 90$. Show that some angle of $x°$ can be constructed with the ruler and dividers iff $\tan x°$ is in $\mathbb{P}$. If V and A are two ruler and dividers points and if $\tan x°$ is in $\mathbb{P}$, then show that there is a ruler and dividers point B such that $m\angle AVB = x$.◊

5.11. Suppose $0 < x < 90$. Show that some angle of $x°$ can be constructed with the ruler and dividers iff both $\sin x°$ and $\cos x°$ are in $\mathbb{P}$.◊

5.12. Sketch a ruler and dividers construction corresponding to Euclid I.1, which asks for an equilateral triangle.◊

5.13. Show that with the ruler and dividers a regular pentagon can be inscribed in a given circle.◊

5.14. Find an iterated pythagorean extension of $\mathbb{Q}$ that contains the field $\mathbb{Q}(\sqrt{31}, \sqrt{43})$.

5.15. If G and H are iterated pythagorean extensions of field F, then show that there is an iterated pythagorean extension of G that contains H.◊

5.16. If a,b,x,y are in $\mathbb{Q}$, $a^2 + b^2$ is not a square in $\mathbb{Q}$, $x + y\sqrt{a^2 + b^2} > 0$, but $x < 0$, then show that $\sqrt{x + y\sqrt{a^2 + b^2}}$ is in $\mathbb{E}$ but not in $\mathbb{P}$. Show $\sqrt[4]{2}$ is in $\mathbb{E}$ but not in $\mathbb{P}$.◊

5.17. Show that $\mathbb{P}(\sqrt[4]{2}) = \mathbb{P}(\sqrt{1 + \sqrt{2}})$, and find a pythagorean extension of $\mathbb{P}(\sqrt[4]{2})$.◊

5.18. What can be done with a ruler and an angle doubler?◊

5.19. Show that $\tan 3°$ is in $\mathbb{P}$. Is $\tan(360/17)°$ in $\mathbb{P}$?◊

5.20. Prove Steiner's ruler and compass construction for Malfatti's Problem: Given $\triangle ABC$ let I be the incenter of the triangle. Then $\overleftrightarrow{IA}$ is one of the two common tangents of the incircles of $\triangle ICA$ and $\triangle IBA$ that intersect $\overline{BC}$; let a be the other. Also, $\overleftrightarrow{IB}$ is one of the two common tangents of the incircles of $\triangle IAB$ and $\triangle ICB$ that intersects $\overline{CA}$; let b be the other. Further, $\overleftrightarrow{IC}$ is one of the two common tangents of the incircles of $\triangle IBC$ and $\triangle IAC$ that intersects $\overline{AB}$; let c be the other. The desired circle with center on $\overleftrightarrow{IA}$ is tangent to $\overleftrightarrow{AB}$, $\overleftrightarrow{AC}$, b, and c; the desired circle with center on $\overleftrightarrow{IB}$ is tangent to $\overleftrightarrow{BC}$, $\overleftrightarrow{BA}$, c, and a; and the desired circle with center on $\overleftrightarrow{IC}$ is tangent to $\overleftrightarrow{CA}$, $\overleftrightarrow{CB}$, a, and b.

6

The Poncelet–Steiner Theorem and Double Rulers

> It is a very different matter actually to carry out the constructions, i.e., with the instruments in hand, than to carry them through, if I may use the expression, simply by means of the tongue.
>
> STEINER

In the early nineteenth century, Poncelet and Steiner dominated the revival of interest in pure geometry, as opposed to the methods of analytic geometry. Jacob Steiner (1796–1863) worked on his father's farm until the age of nineteen, before going off to Berlin to become what some regard as the greatest geometer of modern times. Jean-Victor Poncelet (1788–1867) entered the French army corps of engineers just in time to take part in Napoleon's disastrous 1812 campaign. After his capture by the Russians, Poncelet spent his time in a Moscow prison to good advantage, developing the concepts of projective geometry. Steiner also made significant contributions to this new method of geometric thinking. In 1822, Poncelet, inspired by the results of Mascheroni, gave indications of a proof that all the ruler and compass constructions could be carried out with the ruler alone, provided one circle with its center was given. Steiner published his detailed proof of this result in 1833. Both Poncelet and Steiner were ardent supporters of synthetic geometry and disliked analytic methods to the extent of attacking those who used them. Therefore, it is with apologies to both Poncelet and Steiner that we will use analytic geometry in proving the the-

G. E. Martin, *Geometric Constructions*

orem that bears both their names. However, at this point, we would not consider doing it any other way.

Suppose a transparent sphere, sitting on a plane with its "south pole" the point of contact, has a point light source at its north pole. It can be shown that the shadow on the plane of any circle drawn on the southern hemisphere of the sphere is a circle. (From any point source of light, the shadow on a plane of a circle is always a conic section.) However, the shadow of the center of the circle in the plane of the circle does not generally fall on the center of the shadow circle. From this observation and an argument like that at the beginning of the ruler theory in Chapter 4, we understand why the center of the circle must be either given or obtainable from a starter set in the "constructions with ruler and given circle with its center." Analogous to the "constructions with ruler and given square" of Chapter 4, here the circle is "given" only in the sense that points of the intersection of the circle with a constructed line are taken to be constructed points. We do not suppose all the points of any circle or even of any line are given to us as constructed points. In that case, there would be nothing for us to do as it would follow that all the points in the cartesian plane are constructible. We suppose here that the unit circle with center at the origin has been given to us in the sense just noted. Even so, we need to be specific about a starter set to get underway. Although you may prefer a different starter set, the rest of the following definition should contain no surprises.

Definition 6.1. In the cartesian plane, a point is a ***ruler and circle point*** if the point is the last of a finite sequence $P_1,P_2, \ldots ,P_n$ of points such that each point is in $\{(0,0),\ (2,0),\ (0,2)\}$ or is obtained in one of the two ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; and (ii) as a point of intersection of such a line and the circle through $(1,0)$ with center $(0,0)$. A ***ruler and circle line*** is a line that passes through two ruler and circle points.

It follows immediately that $(1,0)$ and $(0,1)$ are ruler and circle points. The points in the starter set for our ruler theory are ruler and circle points. All the ruler points are ruler and circle points. Further, we have the constructions from the ruler and dividers theory because the given circle permits the constructions of the cannon and more. Instead of determining only the intersection of the unit circle and lines through the origin, here we have the intersection of the unit circle and any line. This turns out to make a big difference, as our next theorem proclaims.

Theorem 6.2 (The Poncelet–Steiner Theorem). *A point is a ruler and circle point iff the point is a ruler and compass point.*

Proof. From the ruler theory, we know that the coordinates of all the ruler and circle points form some field F and that x and y in F implies (x, y)

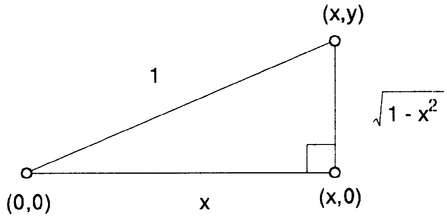


FIGURE 6.1.

is a ruler and compass point. Obviously, $F \subseteq \mathbb{E}$. From Figure 6.1, we see that F has the property that x in F and $-1 < x < 1$ implies $\sqrt{1-x^2}$ is in F. This fact is not as insignificant as it looks at first. Since the fraction $(z-1)/(z+1)$ ranges between -1 and $+1$ for positive values of z, the identity

$$\sqrt{z} = \sqrt{\left(\frac{z+1}{2}\right)^2 - \left(\frac{z-1}{2}\right)^2} = \left(\frac{z+1}{2}\right)\sqrt{1 - \left(\frac{z-1}{z+1}\right)^2}$$

then shows that z in F with $z > 0$ implies $\sqrt{z}$ is in F. So F is a euclidean field contained in $\mathbb{E}$. Therefore, $F = \mathbb{E}$. ∎

Although it is a safe bet that anyone reading this book enjoys an elegant synthetic proof (*i.e.*, one without reference to the coordinate plane), the power of analytic methods is demonstrated by the proof above. In the mathematical literature, a construction with ruler and given circle with its center is usually called a *steiner construction.* In a minor attempt to balance this against the fact that it was Poncelet who first announced the theory, we will refer to the given circle with its center in a steiner construction as the *poncelet circle.* Theorem 6.2 does not really give us any steiner constructions but only informs us what constructions are possible. The exercises will indicate some of the steiner constructions.

Within the proof above of the Poncelet–Steiner Theorem there is a little algebraic result that we state next as a lemma for easy reference in the future.

Lemma 6.3. *Suppose F is a field such that x in F implies $\sqrt{1-x^2}$ is in F when $0 < x < 1$. Then F is a euclidean field.*

Poncelet's famous 1822 work, which reports mainly on the great geometric achievements he made while imprisoned in Russia, includes not only the Poncelet–Steiner Theorem but also the results on the double-edged rulers that we will study in the remainder of this chapter. The first of these is the parallel-ruler, which is simply a ruler with two parallel edges. The width of the parallel-ruler is taken as the unit length. Most rulers that you buy are parallel-rulers; you can easily make one of stiff paper. Of course, such real parallel-rulers have edges of finite length. You may want to pause and

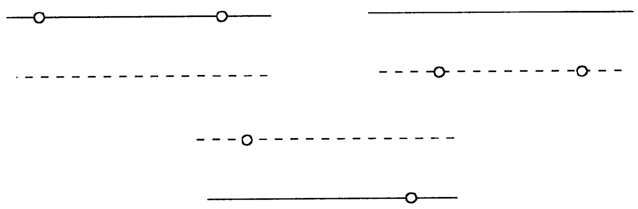


FIGURE 6.2.

play around with a parallel-ruler to see if you can determine the uses of the parallel-ruler. In addition to drawing the line through two given points, we can draw the parallels that are a unit distance from any given line. There is a third use of the parallel-ruler. If the distance between two given points is at least a unit, then we can insert the parallel-ruler between the points such that each edge of the parallel-ruler passes through one of the given points. In this case, an edge of the parallel-ruler determines the line that passes through one of the given points and that is a unit distance from the second point. Another way of looking at this third use is that the parallel-ruler constructs the line through one of the given points and tangent to the unit circle having the second given point as its center. The three uses of the parallel-ruler are illustrated in Figure 6.2. Check that these three uses are reflected in the definition below.

Definition 6.4. In the cartesian plane, a point is a ***parallel-ruler point*** if the point is the last of a finite sequence P_1,P_2, ... ,P_n of points such that each point is in $\{(1,0),(0,1)\}$ or is obtained as the intersection of two lines, each of which satisfies one of the three conditions: (i) the line passes through two points that appear earlier in the sequence; (ii) the line is parallel to and at a unit distance from a line that passes through two points that appear earlier in the sequence; or (iii) the line passes through a point that appears earlier in the sequence and is a unit distance from another point that appears earlier in the sequence. A ***parallel-ruler line*** is a line that passes through two parallel-ruler points.

We will prove that all the ruler and compass constructions can be accomplished with the parallel-ruler.

Theorem 6.5. *A point is a parallel-ruler point iff the point is a ruler and compass point.*

Proof. Starting with $(1,0)$, $(0,1)$, and a parallel-ruler, we can construct, in turn, the six lines having the equations $Y=0$, $Y=1$, $X=0$, $X=1$, $X=2$, and $Y=2$, where the first four of these lines come from the third use of the parallel-ruler. The points in the starter set for the ruler theory are parallel-ruler points. From the ruler theory, we know that the coordinates of all the parallel-ruler points form a field F and that x and y in F implies (x,y) is a parallel-ruler point. A sequence of points satisfying the condition

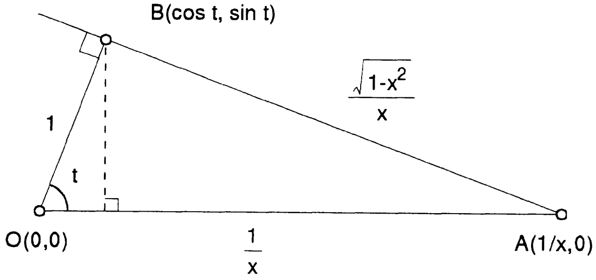


FIGURE 6.3.

of Definition 6.4 must be a sequence of ruler and compass points, since $(1,0)$ and $(0,1)$ are ruler and compass points and since each of the lines in (i), (ii), and (iii) from the definition is easily constructed with the ruler and compass. So $F \subseteq \mathbb{E}$.

Suppose x is in F with $0 < x < 1$. Then $1/x$ is in F and $1/x > 1$. Let $O = (0,0)$ and $A = (1/x, 0)$. The line through A and one unit from O is then a parallel-ruler line. The foot B of the perpendicular from O to this line is a parallel-ruler point by a ruler construction. See Figure 6.3. Since $B = (x, \sqrt{1-x^2})$, then it follows that x in F with $0 < x < 1$ implies $\sqrt{1-x^2}$ is in F. So $\mathbb{E} \subseteq F$ by Lemma 6.3. Hence, $F = \mathbb{E}$, as desired. ∎

All the ruler and compass construction problems can be accomplished with the parallel-ruler. It is fun to try your luck at finding good parallel-ruler constructions. Figure 6.4 with $AB > 1$ should bring a couple to mind.

As a line determines the ideal ruler, so a right angle determines the ideal right-angle ruler. However, a sheet of stiff paper cut at right angles will do nicely. Many stationery stores carry plastic right angles. A right-angle ruler does more than replace the ruler and eliminate the tedious ruler constructions for perpendiculars to a given line through a given point. Placing the right-angle ruler so that each side passes through one and only one of two given points, the vertex lies on the circle having the given points as endpoints of a diameter. The Theorem of Thales is therefore the basis for part of the next definition. We take a starter set that is as small as reasonably possible.

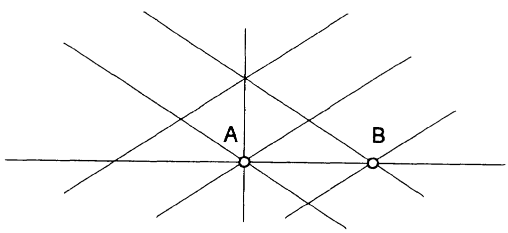


FIGURE 6.4.

Definition 6.6. In the cartesian plane, a point is a ***right-angle point*** if the point is the last of a finite sequence $P_1, P_2, \ldots, P_n$ of points such that each point is in $\{(0,0), (1,0), (0,1)\}$ or is obtained in one of two ways: (i) as the intersection of two lines, each of which either passes through two points that appear earlier in the sequence or is perpendicular to such a line and passes through a point that appears earlier in the sequence; or (ii) as a point of intersection of a line through two points that appear earlier in the sequence and of a circle with a diameter having its endpoints appear earlier in the sequence. A ***right-angle line*** is a line that passes through two right-angle points.

Theorem 6.7. *A point is a right-angle point iff the point is a ruler and compass point.*

Proof. Exercise 6.13. ∎

The same game can be played with angle rulers whose angle is not a right angle. The case for obtuse-angle rulers is only slightly more complicated than that for acute-angle rulers, which we will discuss next. We suppose we have an angle ruler where the angle is acute and has degree measure t where $\tan t^\circ$ is in $\mathbb{E}$. It is even reasonable to suppose the angle has degree measure 30, 45, or 60, since these are the measures of the acute angles encountered in the drafter's plastic triangles that are commonly sold. Although the plastic angles sold in stores always contain a right angle, we restrict ourselves now to using only one acute angle. This time the double ruler consists of the two sides of the acute angle.

Given $\overline{AB}$, Figure 6.5a shows how to construct the perpendicular bisector p of $\overline{AB}$. Figure 6.5b shows how to construct the image P' of point P reflected in given line l.

Figure 6.5c shows how to find the intersection of any given line l and that arc of the circle of all points X on one side of $\overleftrightarrow{AB}$ such that $m\angle AXB = t$. We do this by placing the vertex of the angle ruler on the line l so that each side of the angle is on one and only one of A and B. In particular, the midpoint M of of the major arc with endpoints A and B can be found by taking l to be the perpendicular bisector of $\overline{AB}$. The circle is determined by the three points A,B,M. The circle is C_A where center C is constructed as the intersection of the perpendicular bisectors of $\overline{AB}$ and $\overline{AM}$.

In order for C_A to be a poncelet circle, we must close the gap in Figure 6.5c. That is, we must find the intersection of C_A with a given line m that intersects the minor arc with endpoints A and B. See Figure 6.5d. We construct B' as the image of B under reflection in $\overleftrightarrow{CA}$. Not only is B' on C_A but the minor arc with endpoints A and B' has the same measure as the minor arc with endpoints A and B. Hence, we simply use the technique of Figure 6.5c with B replaced by B' to close the gap. So C_A is a poncelet circle. We have the power of the ruler and a poncelet circle. Therefore, with the proper definition we will have a nice theorem that follows from

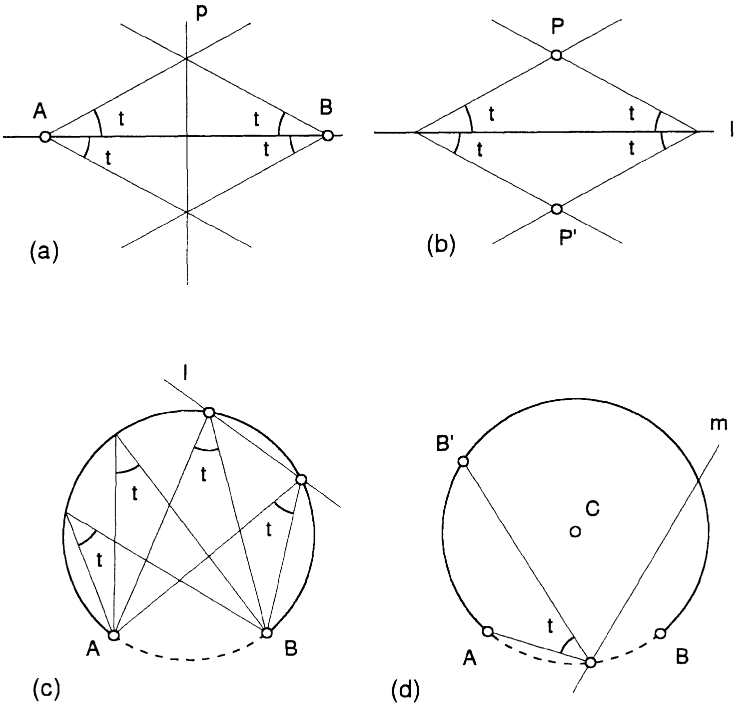


FIGURE 6.5.

the Poncelet–Steiner Theorem. The definition and the theorem are left for Exercise 6.18.

Let's look briefly at adapting the acute-angle ruler constructions of Figure 6.5 to obtuse-angle ruler constructions. There is no problem until we get to Figure 6.5d. This time we have only the minor arc with endpoints A and B, rather than the major arc with the same endpoints, to determine the intersection of the circle containing these arcs and a constructed line. The technique of using reflections still works; only, now we may have to reflect in several radii to get copies of the minor arc to cover the entire circle. It's more work in practice but we end up with the same theory. This idea of using reflections, which can be carried out with the ruler alone, to compensate for a missing arc provides an insight to the Poncelet–Steiner Theorem. It is not necessary that the whole poncelet circle be "given." We need only an arc, however small, to arrive at all ruler and compass constructions.

Exercises

6.1. Sketch a steiner construction that uses the poncelet circle to get a bisected segment on a given line l.◊

6.2. Outline a steiner construction for the angle bisector of a given angle, assuming the analogue of Euclid I.31.◊

6.3. Give a steiner construction for inscribing a square in the poncelet circle.◊

6.4. Given poncelet circle O_U and points A and B with $B \neq 0$, give a steiner construction for the point X on $\overrightarrow{OB}$ such that $OX = OA$.◊

6.5. Outline a steiner construction for the perpendicular to given line l through given point P.◊

6.6. State the steps in a steiner construction for the analogue of Euclid I.3.◊

6.7. Exercise 6.4 leads to a steiner construction for finding the points of intersection of O_A and $\overleftrightarrow{OB}$, where O is the center of the steiner circle. Outline a steiner construction for the intersection of O_A and $\overleftrightarrow{CD}$, when the center O of the steiner circle is off $\overleftrightarrow{CD}$. Outline a steiner construction for the intersection of P_Q and $\overleftrightarrow{RS}$, given points P,Q,R,S.◊

6.8. Sketch a parallel-ruler construction for the analogue of Euclid I.31.◊

6.9. Sketch a parallel-ruler construction for bisecting an angle; sketch a parallel-ruler construction for doubling an acute angle.◊

6.10. Sketch a parallel-ruler construction for a perpendicular to a given line.◊

6.11. Sketch a parallel-ruler construction for bisecting a segment; sketch a parallel-ruler construction for doubling a segment.◊

6.12. Given only a pencil, a standard sheet of paper, and a circular disc, say the top of a can, how can you find the center of the disc.◊

6.13. Prove Theorem 6.7.◊

6.14. Sketch a right-angle ruler construction for bisecting a segment; sketch a right-angle ruler construction for doubling a segment.◊

6.15. Sketch a right-angle ruler construction for bisecting an angle, assuming you can double a segment.◊

6.16. If E is off $\overleftrightarrow{CD}$, give a right-angle ruler construction for the point X on $\overrightarrow{CD}$ such that $CX = CE$.◊

6.17. If E is off $\overleftrightarrow{CD}$, give an acute-angle ruler construction for the point X on $\overrightarrow{CD}$ such that $CX = CE$.◊

6.18. Give a definition of an *acute-angle point* analogous to Definition 6.6. Then state and prove a theorem analogous to Theorem 6.7.

6.19. Give a definition of a *square point*, where the definition models the points constructed by using the construction tool that is a plastic square. Suppose, to make things nice, that the square has sides of length 2. Show that a point is a square point iff the point is a ruler and compass point.

6.20. Give a ruler and compass construction for what is known as *Steiner's Problem*: Given lines l,m,n and points P,Q,R, construct $\triangle ABC$ such that vertices A,B,C are on l,m,n, respectively, and such that extended sides $\overleftrightarrow{BC}$, $\overleftrightarrow{CA}$, $\overleftrightarrow{AB}$ pass through P,Q,R, respectively.

7
The Ruler and Rusty Compass

> Shewing How (Without Compasses) having only a common Meat-Fork (or such like Instrument, which will neither open wider, nor shut closer), and a Plain Ruler, to perform many pleasant and delightful Geometrical Operations.
>
> *Pleasure with Profit*
> WILLIAM LEYBOURN, 1694

A challenging problem is to construct with the ruler and rusty compass a triangle whose sides are congruent to three given segments. It has recently been discovered that the study of the rusty compass goes back to the ancient Greeks. However, the name most associated with the rusty compass is that of the tenth-century Arab scholar Abûl-Wefâ (940–998). Not only did he give the most elementary ruler and rusty compass constructions, but Abûl-Wefâ also gave ruler and rusty compass constructions for inscribing in a given circle a regular pentagon, a regular octagon, and a regular decagon. Ludovico Ferrari (1522–1565) showed in 1547 that all the constructions in Euclid's *Elements* could be done with the ruler and rusty compass. In 1673, Georg Mohr wrote a little book that also proves the principal theorem of this chapter.

You have been following the development up to this point. Now, you are asked to write the rest of this chapter yourself. That's right; this is a do-it-yourself chapter. You have all the information you need to give a good definition and then produce a fine theorem. Of course, you will want to do

G. E. Martin, *Geometric Constructions*

some ruler and rusty compass constructions and make up some exercises too. You may want to leave most of these tasks for the exercises, but your exercises should match the hints and answers in The Back of The Book.

Good luck!

8
Sticks

"and you are as stiff as a stick. Dull. Conventional and repressed. Tied and trammeled. Subdued, smothered and stifled. Squashed, squelched and quenched." [So said the dot.]
"Why take chances," replied the line without much conviction. "I'm dependable. I know where I'm going. I've got dignity!"

The Dot and the Line
NORTON JESTER

Our task is to form a geometric construction theory that models what we can do with an inexhaustible supply of toothpicks. It should be clear that in order to do the modeling, we must first have some familiarity with what it is we are to model. Therefore, we should first play around with stacks of toothpicks just to see what we can do with them. Paper and pencil is just not satisfactory; straws or even long, thin strips of paper will serve the purpose. Reading about toothpick constructions is not the same thing as discovering for yourself some of the things that follow. Since the results of this chapter are not used later, you can go on to the next chapter now and read this chapter after you have had a chance to play with toothpicks.

In case you need some motivation or didn't get too far playing with toothpicks, consider the following puzzle: ***Starting with a given toothpick and a point in its interior, build a square having the toothpick as***

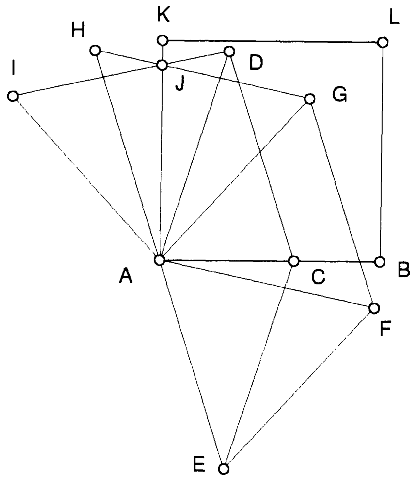


FIGURE 8.1.

a side. To make the puzzle more interesting, here we impose the handicap that two toothpicks are not allowed to overlap at more than one point. Perhaps you should convince yourself first that this self-imposed handicap is no hindrance. The construction of the square is not easy and can be done with a total of sixteen toothpicks. The puzzle is solved in Figure 8.1, where the given points are A,B,C and the other points are determined in alphabetical order. If you solve the puzzle or even if you study the solution to fully understand the construction, you will have mastered almost all the principal ideas in the theory of toothpick constructions. We will refer back to this solution of the puzzle several times. If you need a hint, consider that the angle bisectors of a linear pair of angles are perpendicular. (If P–Q–R and S is off $\overleftrightarrow{PR}$, then the angle bisectors of $\angle PQS$ and $\angle RQS$ are perpendicular.) Stop peeking at the solution; try the puzzle yourself. We'll wait for you.

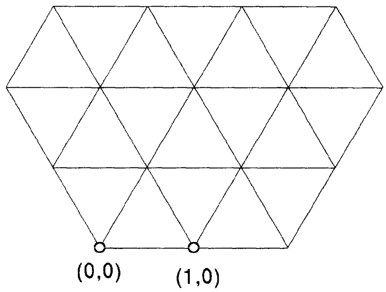


FIGURE 8.2.

There is no question that, if anything, the first thing we do with a pile of toothpicks is to make an equilateral triangle. We then continue to form a triangular grid or triangular lattice like that suggested by Figure 8.2. Staring at this lattice for a while should tell us something. First we have used segments to model the toothpicks. In the model the lattice is infinite; it's a safe bet that your toothpick constructions are finite. Perhaps the ends of the toothpicks you are using are not the same. Since we are not picking our teeth with these (please), this difference is of no concern to us and we ignore the difference in the model. Also, it seems that for our interest we may as well suppose the toothpicks have no thickness. The triangular lattice is uniquely determined, once one toothpick is put down. However, in order to put down one toothpick, we should have two points given so that we can put one end of a toothpick on one point and have the toothpick pass through the other point. As usual, we do not want to leave anything to chance. Of course we suppose $(0,0)$ is one of the given points and that the length of a toothpick is 1. Did you notice that we are assuming we are working in the plane and not space? More formally, we are saying the cartesian plane models our desk top. If the second given point were $(5,0)$, we would be out of luck as we could not put down any toothpicks. Even two toothpicks, one extended from each of the points, would not reach between $(0,0)$ and $(5,0)$. We certainly do not allow ourselves to put a toothpick down to "aim" at some point. A better selection for the second given point is $(1,0)$. Better, yes, but not a best selection. With $(1,0)$ as the second given point, we are essentially done because we can place no more toothpicks than those in the lattice. In the puzzle, we had to be given a point C on "toothpick-AB" or otherwise we could have done nothing more than build the triangular lattice on the toothpick-AB. Since we want a more complicated theory than that generated by just the lattice, we suppose the second given point is $(1/2,0)$. These two points will be sufficient for our starter set.

Toothpicks are sticks and we are going to start calling them that here. This arbitrary change in language is mainly for aesthetic reasons. Later we shall be using the terms "stick point" and "stick line," which are much more flowing than the awkward alternative terms "toothpick point" and "toothpick line," especially when encountered over and over again in one paragraph. This change is made for easier reading and smoother speech. Further, the terminology is closer to that of T. R. Dawson, the English mathematician who first described these constructions in his 1939 paper *"Match-Stick" Geometry.*

We now wish to make precise how we may use sticks to construct points. Figure 8.3 illustrates the methods. Initially, we must be given two points in the beginning in order to get started. We have already suggested the method of Figure 8.3b, where we put a stick down with one end at a given point A to pass through a given point B in order to determine the point P at the other end of the stick. We have also noted a special case of the method

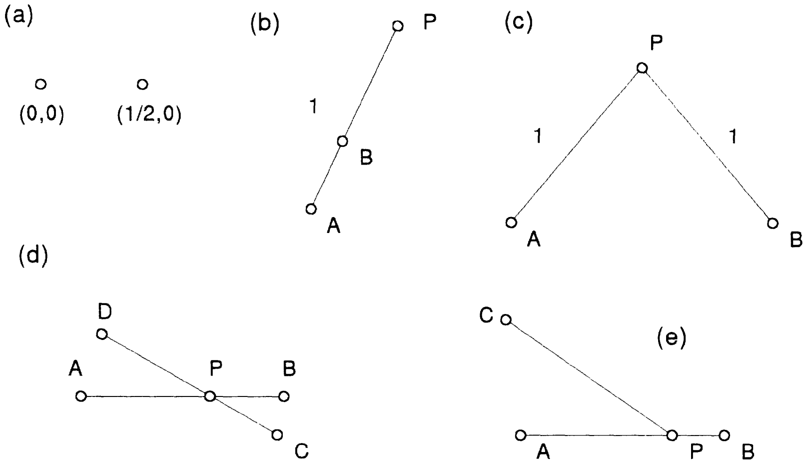


FIGURE 8.3.

of Figure 8.3c, where we form an isosceles triangle having sticks as at least two of its sides when given a short enough base. In particular, when the given base is a stick, we have the special case of the isosceles triangle being equilateral. Of course, as in Figure 8.3d, the intersection of two distinct sticks not on one line should determine a point. The last method is shown in Figure 8.3e. If we can rotate a stick about one end so that the other end is the only point of intersection with a second given stick, then the point of intersection is considered to be constructed. In other words, if point P is in the intersection of a stick and a unit circle whose center is a given point off the line containing the stick, then P is constructible.

The following definition allows us to distinguish by name those points and lines we can construct with sticks. The definition of a stick point follows the previous paragraph and Figure 8.3. A stick models our original toothpick.

Definition 8.1. In the cartesian plane, a point is a ***stick point*** if the point is the last of a finite sequence $P_1, P_2, \ldots, P_n$ of points such that each point is in $\{(0,0),\ (1/2,0)\}$ or is obtained in one of four ways: (i) as the point P on $\overrightarrow{AB}$ such that $AP = 1$ where A and B are two points that appear earlier in the sequence and $AB < 1$; (ii) as a point that is one unit from both of two points A and B that appear earlier in the sequence and $AB < 2$; (iii) as the only point of intersection of two unit segments each of which has endpoints that appear earlier in the sequence; and (iv) as the intersection of a unit segment whose endpoints appear earlier in the sequence and of a unit circle having a center that appears earlier in the sequence and off the line containing the unit segment. A ***stick*** is a segment $\overline{PQ}$ where P and Q are stick points with $PQ = 1$. A ***stick line*** is a line that contains a stick.

Check that the definition gives us the expected theorem that describes the basic stick constructions.

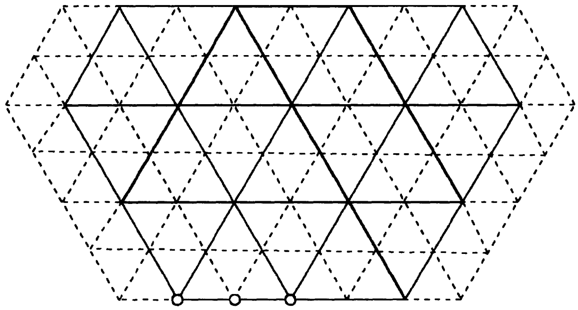


FIGURE 8.4.

Theorem 8.2. *If A and B are two stick points and $AB < 1$, then the point P on $\overrightarrow{AB}$ such that $AP = 1$ is a stick point. If A and B are stick points and $AB < 2$, then the two points that are one unit from both A and B are stick points. The point of intersection of two intersecting but noncollinear sticks is a stick point. A point of intersection of a stick and a unit circle having a stick point as center is a stick point.*

You may be surprised at the definition of a stick line. We could have followed the format of our previous definitions and defined a line through two stick points to be a stick line. Note, we did not do this. In either case, we would have to prove eventually that a line through two stick points contains a stick. A choice had to be made, and the path chosen seems more natural when toothpicks are in hand. The choice emphasizes that we do not yet have a ruler. So, a stick line is a line that contains a stick by definition. It is not obvious that two stick points determine a stick line when the points are far apart. Exercises 8.1, 8.2, and 8.3 concern aspects of Definition 8.1. You may want to take a look at these three exercises at this time. Although you cannot do the first exercise now, you may want to be thinking about it as you go along. In particular, read Exercise 8.3 now if you feel something important has been left out of the definition.

We now suppose we have put down the first stick with one end at $(0, 0)$ and, since the stick passes through $(1/2, 0)$, the other end is at $(1, 0)$. Then there exists the triangular lattice of Figure 8.2 built on this first stick. If you recall that the sides of a 30-60-90-triangle with hypotenuse 1 are $1/2$ and $\sqrt{3}/2$, then you know the third vertex of the first equilateral triangle formed is $(1/2, \pm\sqrt{3}/2)$. It follows that the points of the lattice are precisely the rusty compass points. By Theorem 3.17, these are the points $(m + (n/2), n\sqrt{3}/2)$ where m and n are integers. Compare Figure 8.2 with Figure 3.8a.

If we place one end of a stick at $(1/2, 0)$ and have the stick pass through $(0, 0)$, then the triangular lattice determined by this stick is congruent to that of Figure 8.2, only translated. The two lattices together determine

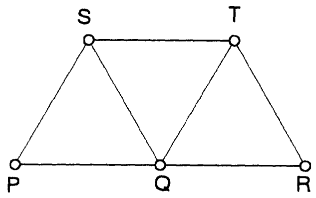


FIGURE 8.5.

another triangular lattice, shown in Figure 8.4. The points of this third lattice are those with coordinates of the form $((m/2)+(n/4), n\sqrt{3}/4)$ where m and n are integers. At present, we know points $(1/4, \sqrt{3}/4)$ and $(9, 5\sqrt{3})$ are stick points; we do not yet know whether either of the points $(1/3, 0)$ and $(0, \sqrt{7})$ is a stick point or not; and we might suspect point $(\pi, 0)$ is not a stick point. Several results are apparent from looking at these lattices. The first is singled out for special emphasis.

Theorem 8.3 (The Extension Theorem). *If $\overline{PQ}$ is a stick and Q is the midpoint of $\overline{PR}$, then $\overline{QR}$ is a stick. Every point on a stick line is on a stick that is on that line. The intersection of two nonparallel stick lines is a stick point.*

The second and third parts of the Extension Theorem follow directly from the first, and the first part is trivial, unless you have been only reading about toothpicks and not actually playing with them. Surely among the first things to do with toothpicks is to create Figure 8.5. This is part of the first triangular lattice; starting with stick $\overline{PQ}$, construct equilateral triangles $\triangle QPS$, $\triangle QST$, and $\triangle QTR$, in turn, to determine stick $\overline{QR}$ and prove the theorem. Note that the conclusion of the first part of the Extension Theorem could be rephrased to state that R is a stick point. In the puzzle of Figure 8.1, the reason for sticks $\overline{EF}, \overline{AF}, \overline{FG}, \overline{AG}, \overline{GH}$ is to determine stick $\overline{AH}$, which "extends" segment $\overline{AE}$. If you have not mastered the puzzle, now take away sticks $\overline{AF}, \overline{EF}, \overline{FG}$ for a different view.

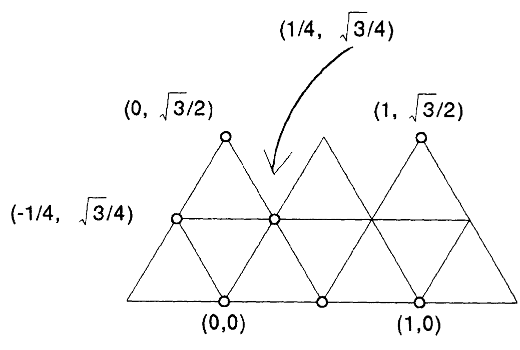


FIGURE 8.6.

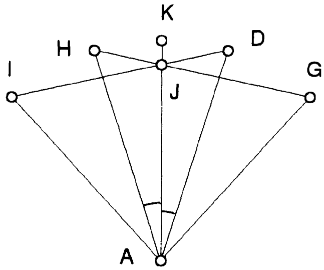


FIGURE 8.7.

Figure 8.5 is a detail from the triangular lattice of Figure 8.2; Figure 8.6 is a detail from the triangular lattice of Figure 8.4. The following theorem should be evident from these triangular lattices.

Theorem 8.4. *A line parallel to a stick line at a distance* $\sqrt{3}/2$ *from that line is a stick line. Each stick point is within* $1/2$ *of three other noncollinear stick points. Each stick* $\overline{AB}$ *contains a stick point between* A *and* B. *If* p *and* q *are integers, then* (p,q) *is a stick point.*

The first statement in Theorem 8.4 rephrases the following: If $\overline{AB}$ is a stick, then each of the two lines parallel and at a distance $\sqrt{3}/2$ from $\overleftrightarrow{AB}$ contains a stick. This should be clear from Figure 8.5, where $\overleftrightarrow{ST}$ is $\sqrt{3}/2$ from $\overleftrightarrow{PQ}$. Looking at the triangular lattice in Figure 8.4, we see that not only is every stick point within $1/2$ of three noncollinear stick points, so is every point in the plane. That the interior of any stick must contain a stick point follows from the fact that you cannot place a stick on the triangular lattice of Figure 8.4 without the interior of the stick intersecting a line of the lattice different from the line containing that stick. From the lattice of Figures 8.4 and 8.6, we see that for integer p the points $(p,0)$ and $(p,\sqrt{3}/2)$ are stick points. Since $\sqrt{3}/2 < 1$, then the stick with one end $(p,0)$ that passes through $(p,\sqrt{3}/2)$ has its other end at $(p,1)$. So $(p,1)$ is a stick point. Starting with the stick with endpoints $(p,0)$ and $(p,1)$, we see that it follows that (p,q) is a stick point for integer q by at most $|q|$ applications of the Extension Theorem. All of Theorem 8.4 should now be clear.

Did a different proof that (p,q) is a stick point for integers p and q occur to you? In the puzzle, a unit square is constructed on a given stick. Continuing to build squares on the sides of already constructed squares, we obtain in the model the unique square lattice based on the stick with endpoints $(0,0)$ and $(1,0)$. In other words, all points (p,q) with p and q integers are points of the square lattice and so are stick points. The square lattice reminds us of standard graph paper. So the puzzle, whose solution implicitly contains the Extension Theorem, also shows that (p,q) is a stick point for integers p and q.

Let's analyze the puzzle further. Starting with $\overline{AC}$ on stick $\overline{AB}$, we build isosceles triangles to get stick points D and E. These points are symmetric

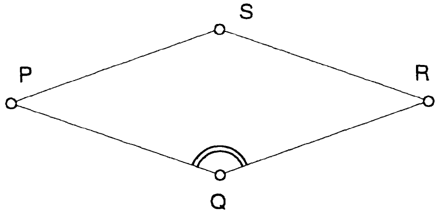


FIGURE 8.8.

with respect to $\overleftrightarrow{AB}$. So the stick $\overline{AB}$ bisects $\angle DAE$. With the Extension Theorem we get stick $\overline{AH}$. Since $\angle HAD$ and $\angle DAE$ form a linear pair of supplementary angles, their angle bisectors are perpendicular. (The sum of half of each of two numbers that sum to 180 must be 90.) So the problem now is to bisect $\angle HAD$. Part of Figure 8.1 is reproduced in Figure 8.7. Stick points H and D are, of course, symmetric with respect to the sought-after angle bisector; the stick points G and I are constructed to be symmetric with respect to the angle bisector also. Hence the sticks $\overline{GH}$ and $\overline{DI}$ are symmetric to the angle bisector, and, since these sticks intersect at stick point J, then J must sit on the angle bisector. In other words, $\overrightarrow{AJ}$ is the angle bisector. The point K is then easily determined such that $\overline{AK}$ is a stick perpendicular to stick $\overline{AB}$. The puzzle is then solved with the one more move that consists of building the sides of an isosceles triangle having base $\overline{BK}$. Not only have we constructed a perpendicular and a square, we have encountered a means of bisecting angles. Although the construction of Figure 8.7 can be adapted to angles greater than 120°, the one step construction of Figure 8.8 is nicer. Here, sticks $\overline{QP}$ and $\overline{QR}$ determine the sticks $\overline{SP}$ and $\overline{SR}$ such that $\square PQRS$ is a rhombus. Since $m\angle PQR \geq 120$, then $QS \leq 1$ and $\overrightarrow{QS}$ bisects $\angle PQR$. There is one fly in the ointment, however.

From the puzzle, we know we can erect a perpendicular at a given point. Now suppose $\overline{PQ}$ and $\overline{QR}$ are sticks forming $\angle PQR$. If $\angle PQR$ is at least 120°, the method of Figure 8.8 uses three sticks to get a stick on the angle bisector quickly. If $\angle PQR$ is less than 120°, then the method of Figure 8.7 uses five sticks to get the angle bisector in all cases but one. If the given angle is 60°, then the construction in Figure 8 7 collapses; in Figure 8.7 the sticks $\overline{GH}$ and $\overline{DI}$ coincide and no longer determine point J. It is what we might expect to be the easy case of a 60° angle that gives all the trouble. There are several ways out of this dilemma, however. We can use the fact that $60 + 60 - 90 = 30$ and that we know how to erect a perpendicular at a point. This inelegant approach is illustrated in Figure 8.9, where $\angle PQR$ is the given angle of 60°, $\triangle QRS$ is equilateral, and $\angle SQT$ is right. Nevertheless, we now know we can bisect any given angle and have therefore completed a proof of the next theorem.

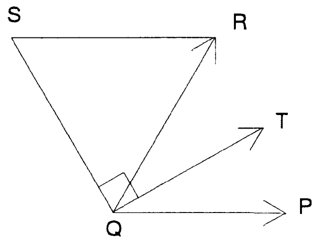


FIGURE 8.9.

Theorem 8.5. *The perpendicular to a stick line at a stick point is a stick line. The angle bisectors of the angles formed by two stick lines are stick lines.*

Closely related to bisecting the angle determined by two sticks is the problem of bisecting short segments. For example, in Figure 8.10, $\overline{AB}$ with $AB < 1$ is bisected by constructing isosceles triangle $\triangle ADB$ and then bisecting the angle at D, the stick $\overline{PQ}$ can be bisected with our inelegant bisection of $\angle PRQ$, and the segment $\overline{UV}$ with $1 < UV < 2$ is bisected by the intersection of the angle bisectors of $\angle UWV$ and $\angle WUX$ of unit rhombus $\square UWVX$.

Another method of bisecting a stick, and therefore of bisecting an angle of $60°$, is to form a triangle in which only one side is a stick and the other sides are shorter than a stick. Now it is easy to bisect the two shorter sides and then use the fact that the medians of a triangle are concurrent to find the midpoint of the third side. This construction takes a lot of sticks, however. A more elegant construction for bisecting stick $\overline{AB}$ and that uses only thirteen toothpicks goes as follows. Let $\overline{BC}$ be any stick such that $\angle ABC$ is less than $60°$. By building an isosceles triangle with sides of unit length on base $\overline{AC}$, construct D such that $\square ABCD$ is a unit rhombus. Extend stick $\overline{AD}$ by the construction of stick $\overline{AG}$ as in the Extension Theorem and as in Figure 8.11. (Point E is taken on the same side of $\overleftrightarrow{AD}$ as B so that $\triangle AFG$ can be used again.) Since $\overline{BC}$ is parallel and congruent to $\overline{AD}$, then $\overline{BC}$ is parallel and congruent to $\overline{GA}$. So $\square BCAG$ is a parallelogram having $\overline{AB}$ as a diagonal. Hence, the equilateral triangles $\triangle BHC$ and $\triangle AFG$ are symmetric about M, the midpoint of $\overline{AB}$. Therefore, a stick covering $\overline{IJ}$ in

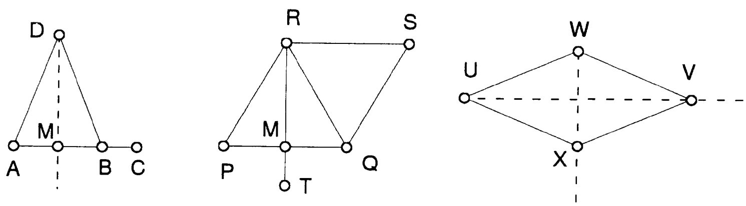


FIGURE 8.10.

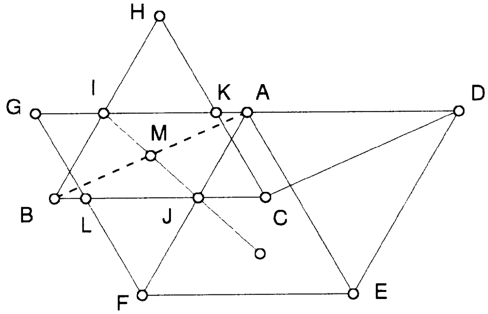


FIGURE 8.11.

the figure will intersect $\overline{AB}$ at M, as desired. A curiosity of this construction is that the midpoint of stick $\overline{AB}$ can be determined with thirteen toothpicks excluding $\overline{AB}$, since M is the intersection of a stick covering $\overline{IJ}$ and a stick covering $\overline{KL}$ of Figure 8.11. Exercise 8.7 shows that a stick can be bisected with a construction that is based on the Second Push-up-and-Pull-down Theorem; this remarkable construction uses only eleven sticks.

When we face the problem of bisecting segments of length at least 2, we come face to face with a larger problem. We do not now have a ruler in our construction theory. If a line contains a stick, then we can build as much of the line as we want by the Extension Theorem. However, if A and B are stick points with, say, $AB = 17$, then it is not clear that we can bridge the gap from A to B with sticks. We would like to show that a line containing two stick points is a stick line. At present, we do not know whether two lines, each of which contains two stick points, intersect at a stick point or not. Our problems will be resolved once we can show the midpoint of two stick points is a stick point. To do this we need one more theorem.

Theorem 8.6 (*cf.* Euclid I.31). *The line through a stick point that is parallel to a stick is a stick line.*

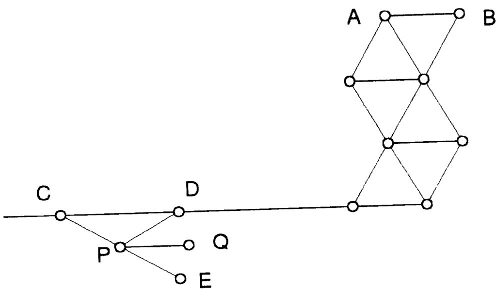


FIGURE 8.12.

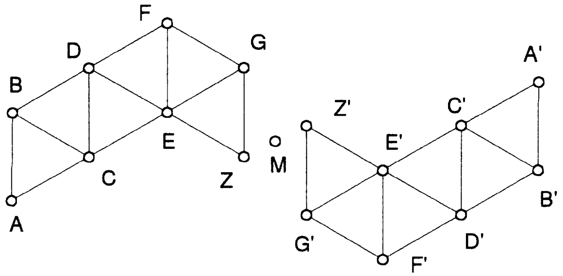


FIGURE 8.13.

Proof. We suppose we are given stick point P and stick $\overline{AB}$. We wish to show that the line through P that is parallel to stick $\overline{AB}$ contains a stick. By several applications of the first part of Theorem 8.4, we may suppose that we have a stick line parallel to stick $\overline{AB}$ and at distance less than 1 from P. See Figure 8.12. The unit circle with center P intersects this line at two stick points C and D. Now if E is such that P is the midpoint of C and E, then $\overline{PE}$ is a stick. (This should now begin to look like part of the solution to the puzzle.) Hence the angle bisector of $\angle EPD$ is a stick line by Theorem 8.5. Since the angle bisector of $\angle CPD$ is perpendicular to $\overleftrightarrow{CD}$ and to the angle bisector of $\angle EPD$, then the angle bisector of $\angle EPD$ is parallel to $\overleftrightarrow{CD}$ and hence to $\overleftrightarrow{AB}$. ■

We are now ready to prove the important theorem about midpoints.

Theorem 8.7 (Stick Midpoint Theorem; *cf.* Euclid I.10).
The midpoint of two stick points is a stick point.

Proof. Suppose A and A' are two stick points with midpoint M. If $AA' < 2$, then we are done by our previous constructions from Figure 8.10. Let $\overline{AB}$ be a stick with B off $\overleftrightarrow{AA'}$; such a point B exists by the second part of Theorem 8.4. There is a point B' such that $\square ABA'B'$ is a parallelogram. See Figure 8.13. Point B' is a stick point by the previous theorem. The parallelogram has M as its point of symmetry. Therefore, the triangular lattice based on stick $\overline{AB}$ and the triangular lattice based on stick $\overline{A'B'}$ are together symmetric about M. It follows that there are stick points Z and Z', one from each of the two lattices and such that their midpoint is M, with $ZZ' < 1$. Since each of Z and Z' is a stick point, then M is a stick point. ■

If A and B are stick points, then by taking successive midpoints we can find a stick point C on $\overleftrightarrow{AB}$ such that $0 < AC < 1$. Then $\overleftrightarrow{AB}$ is a stick line. So the Stick Midpoint Theorem has the corollary that a line is a stick line iff the line contains two stick points.

Theorem 8.8. *The line through two stick points is a stick line.*

Since the line through two stick points is a stick line, since the intersection of two distinct stick lines is a stick point, and since the intersection of a stick line and the unit circle is a stick point, then it follows from the Poncelet–Steiner Theorem that every point constructible by ruler and compass is a stick point. A glance back at our Definition 8.1 will show that every stick point can be constructed by a ruler and compass. Hence we have our final theorem.

Theorem 8.9. *A point is a stick point iff the point is a ruler and compass point.*

The theorem proves all ruler and compass constructions are possible with toothpicks; the theorem does not tell how to do these constructions. Are there shorter constructions than those that have been given above? In particular, is there a more elegant way to bisect an angle of $60°$?

Exercises

8.1. How would our theory have been changed if in Definition 8.1 a *stick line* had been defined as a line containing two stick points? Which definition do you prefer?◊

8.2. How would our theory have been changed if "off the line containing the stick" were deleted from the end of the definition of a stick point in Definition 8.1?◊

8.3. Suppose A and B are given points such that $1 < AB < 2$. Fixing one end of a toothpick at A and one end of another toothpick at B, we can rotate the toothpicks until the two toothpicks overlap. Should we consider the points determined by the ends of the toothpicks as constructed points? In other words, do you think we should have written into Definition 8.1 that the points C and D on $\overline{AB}$ such that $AC = BD = 1$ are stick points when A and B are given stick points such that $1 < AB < 2$?◊

8.4. Outline a second toothpick construction for the analogue of Euclid I.31.◊

8.5. Outline a toothpick construction for the image of stick point P under the reflection in stick line $\overleftrightarrow{AB}$.

8.6. Outline a toothpick construction for the vertices of a regular octagon.

8.7. Using only eleven sticks, illustrate a toothpick construction for bisecting one of the sticks with a given interior point.◊

8.8. If $1 < AB < \sqrt{3}$, sketch a toothpick construction that takes only ten sticks to bisect $\overline{AB}$.◊

8.9. If $\sqrt{3} < AB < 2$, sketch a toothpick construction that takes only ten sticks to bisect $\overline{AB}$.◊

8.10. If A and B are stick points such that $\sqrt{3} < AB < 2$, sketch a toothpick construction that takes only nine sticks to get a stick that is perpendicular to $\overleftrightarrow{AB}$ at A.◊

9
The Marked Ruler

After the dark ages in Europe, the first important mathematical work was that of the Italian school. It contributed to arithmetic and produced the ultimate in classical algebra. Symbolic algebra *per se* first appeared at the very end of the sixteenth century, however, in the ideas of Vieta, who preceded by a short time interval the men whom we have called the forefathers of modern mathematics, namely, Descartes and Fermat.
EDNA E. KRAMER

A marked ruler is simply a straightedge with two marks on its edge. With an appropriate starter set, an immediate use of the marked ruler is to draw the line through two given points and mark off unit segments on the line. Since the two marks thus provide a scale in the sense of a rusty dividers, then the ruler and dividers theory can be applied to marked ruler constructions. Throughout this chapter, the two points determined by the two marks on the marked ruler in a particular position will be called R and S. We suppose the marks are one unit apart. So $RS = 1$, as in Figure 9.1a. For a second use of the marked ruler, we can set one mark on a given point R and rotate the marked ruler until the second mark falls on a given line s at a point S, whenever the unit circle with center R intersects the line s. See Figure 9.1b. The unit circle with center R can be taken as a

G. E. Martin, *Geometric Constructions*

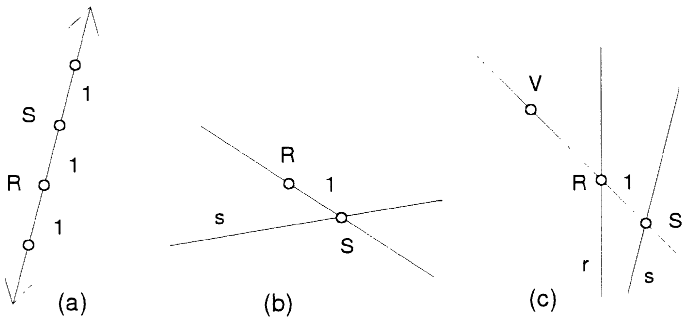


FIGURE 9.1.

poncelet circle. The Poncelet–Steiner Theorem now implies all the ruler and compass constructions are possible with the marked ruler alone. The characteristic use of the marked ruler is called *verging* or *insertion.* Given point V and two lines r and s, by *verging through V with respect to r and s* we determine two points R and S that are one unit apart and such that V is on $\overleftrightarrow{RS}$, R is on r, and S on s. See Figure 9.1c. Thus, in verging through V with respect to r and s, the marked ruler is placed down to pass through V with one mark on r and the other on s. Using the marked ruler in this way to solve two of the three classical construction problems goes back to the Greeks who, according to Pappus, "moved a ruler about a fixed point until by trial the intercept was found to be equal to the given length." A verging is sometimes called by its Greek name *neusis*. Although Apollonius's book *Neusis* on the subject is lost, J. P. Hogendijk has reconstructed the text from Arabic traces of the work.

It is its use for verging that makes the marked ruler such a powerful construction tool. The second mentioned use of the marked ruler, which is illustrated in Figure 9.1b, is actually only the special case of verging where $R = V$ and r is any constructible line through R. Further, marking off units from a point on a line, which is illustrated in Figure 9.1a, is only the special case of the second use where R is on s. Therefore, the two fundamental operations a marked ruler performs are to draw lines through two constructed points and to verge through a constructed point with respect to two constructed lines. Check that Definition 9.1 does model what we can do with the marked ruler and that the definition implies Theorem 9.2.

Definition 9.1. In the cartesian plane, a point is a ***marked ruler point*** if the point is the last of a finite sequence $P_1, P_2, \ldots, P_n$ of points such that each point is in $\{(0,0), (1,0), (0,1)\}$ or is obtained in one of two ways: (i) as the intersection of two lines, each of which passes through two points that appear earlier in the sequence; and (ii) as either of two points that are one unit apart, that are collinear with a point that appears earlier in the sequence, and that are such that each lies on a different line through two

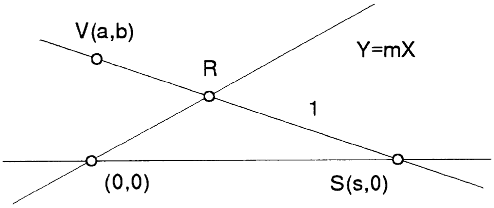


FIGURE 9.2.

points that appear earlier in the sequence. A ***marked ruler line*** is a line that passes through two marked ruler points. A ***marked ruler circle*** is a circle through a marked ruler point with a marked ruler point as center. A number x is a ***marked ruler number*** if $(x, 0)$ is a marked ruler point. Let $\mathbb{V}$ be the set of marked ruler numbers.

Theorem 9.2. *The intersection of two marked ruler lines is a marked ruler point. If points R and S are on different marked ruler lines, are one unit apart, and are collinear with a marked ruler point, then R and S are marked ruler points.*

From the definition and the remarks preceding the definition, it is evident that all steiner constructions are available to us and that, therefore, we can execute any ruler and compass construction on the set of marked ruler points. In particular, it follows from the theory for ruler and compass constructions that $\mathbb{V}$ is a euclidean field. As we might expect, it will turn out that $\mathbb{V}$ is larger than $\mathbb{E}$. Let's look at verging algebraically, with reference to Figure 9.2, where we have verged through point (a, b) with respect to the lines with equations $Y = mX$ and $Y = 0$. We would like to express s in terms of the given numbers a,b,m. Now R has coordinates (x_0, y_0) where $x_0 = bs/(ms - ma + b)$ and $y_0 = mx_0$. Substituting these coordinates into the equation $(x_0 - s)^2 + (y_0 - 0)^2 = (RS)^2 = 1$, we have a fourth degree polynomial in s with coefficients in terms of a,b,m. So verging allows us to solve some quartic equations. This is not a surprise since verging can have as many as four solutions, as in Figure 9.3, where the four ways to verge through V with respect to lines r and s are shown. You should verify that verging through a point with respect to two perpendicular lines or with respect to two parallel lines also gives a polynomial equation of degree at most 4. Therefore, verging produces only points having coordinates that are solutions to polynomial equations with coefficients in $\mathbb{V}$ and of degree at most 4.

We have noted that verging allows us to solve some quartic equations. Which ones? Eventually, we shall show that with the marked ruler we can solve all problems whose algebraic solution depends on solving a cubic or quartic polynomial with coefficients in $\mathbb{V}$. To do this, we first follow the Greeks to solve the two specific problems of trisecting the angle and constructing a cube root. After that, we will show that if we can do these

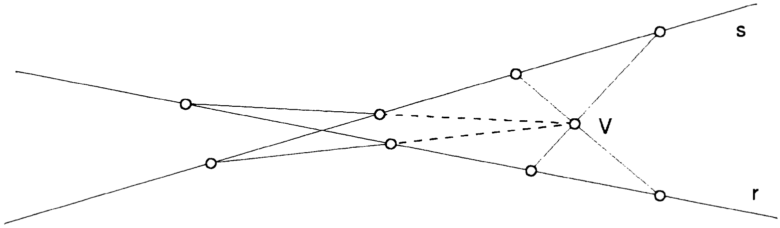


FIGURE 9.3.

two problems, then we can do all the problems dependent on equations of degree at most 4.

François Viète (1540–1603), better known by his Latin pen name Vieta, was the first to use letters to represent unknowns. Letters had been used to represent numbers as far back as Diophantus of Alexandria, the greatest algebraist of antiquity. However, Vieta's *Introduction to the Analytic Art* of 1591 used letters purposefully and systematically as general coefficients, with the consonants for constants and the vowels for unknowns. This practice of using letters for given and unknown quantities is one of the important steps in the development of mathematics. The new language of symbolic algebra is of such importance that Vieta has been called the Prometheus of Mathematics. The introduction of this symbolic language is the watershed that separates modern mathematics from what came before. The only comparable epoch-making event in the history of mathematics is the introduction of deductive reasoning, which is traditionally credited to Thales and cited as the beginning of mathematics as we know it.

Trained as an attorney, Vieta served the courts of Henri III and of his successor Henri IV. As well as being a counselor to individuals, parliament, and kings, Vieta acted as a cryptanalyst. He was so successful in decoding intercepted messages in 1589 that the enemies of Henri III claimed that Vieta's decipherment could have been achieved only by sorcery and necromancy. Another story is more directly related to mathematics. Chided by the ambassador from The Netherlands that France had no mathematician that could solve a certain forty-fifth degree equation, Henri IV called upon Vieta to meet the challenge. Vieta immediately saw that the equation was satisfied by the chord of a unit circle that subtends an angle of radian measure $2\pi/45$ and provided his king with one solution within minutes and twenty-two more solutions the next day. Since the remaining solutions were negative they did not make sense to Vieta. In spite of his aversion to negative solutions to equations, Vieta's work on algebra, the analytic art, changed the look of mathematics.

Vieta declared the Platonic restriction a "defect of geometry" and recommended the remedy that verging be allowed as a new postulate. He was the first to show that if we can solve the two specific problems of trisecting the angle and constructing a cube root then we can do all the problems

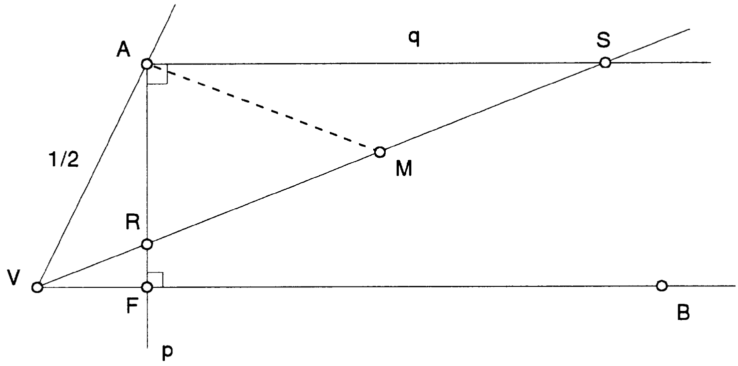


FIGURE 9.4.

dependent on equations of degree at most 4. As Vieta says at the end of his *A Supplement to Geometry* of 1593, This is very worth knowing.

The first of the two specific problems is angle trisection. Our solution is due to Pappus, the last of the giants in Greek mathematics. His greatest work, the *Collection*, was written in Alexandria about AD 320. In addition to his own contributions to mathematics, we learn from this book much about earlier Greek mathematics that we would not otherwise know. Pappus describes some solutions to the three famous problems from antiquity and virtually asserts that the classical problems are impossible under the Platonic restriction. As we know, proofs for this assertion were first given by Wantzel in the nineteenth century. In Figure 9.4, Pappus gets the trisector of $\angle AVB$ by verging through V with respect to the lines p and q.

Theorem 9.3 (Pappus). *Given acute angle $\angle AVB$ with $AV = 1/2$, let $p = \overleftrightarrow{AF}$, where F is the foot of the perpendicular from A to $\overleftrightarrow{VB}$. Let q be the perpendicular to p at A. Let $\overrightarrow{VR}$ intersect p at R between A and F and intersect q at S such that $RS = 1$. Then $\overrightarrow{VR}$ trisects $\angle AVB$.*

Proof. Let $\angle BVS$ have measure t. Then $\angle VSA$ has measure t. Let M be the midpoint of R and S. Since $\angle RAS$ is right, then A lies on the circle with diameter $\overline{RS}$ by the converse of the Theorem of Thales. So $MA = MR = MS = VA = 1/2$, and $\triangle AMS$ and $\triangle MAV$ are isosceles triangles. Hence, $\angle MAS$ has measure t by the Pons Asinorum, $\angle VMA$ has measure $2t$ by the Exterior Angle Theorem, and $\angle AVM$ has measure $2t$ by the Pons Asinorum. Therefore, $\overrightarrow{VR}$ trisects $\angle AVB$. ∎

Note that in Figure 9.4 the perpendicular bisector of $\overline{VS}$ intersects $\overline{AS}$ at a point T such that $\overrightarrow{VT}$ is the other angle trisector of $\angle AVB$. The construction for the trisectors works for obtuse angles as well, provided R is taken such that A–F–R.

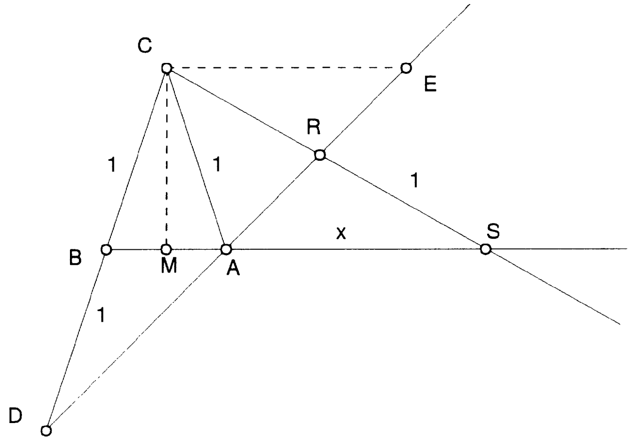


FIGURE 9.5.

Corollary 9.4 (Trisection Theorem). *If* $\cos x$ *is a marked ruler number, then* $\cos(x/3)$ *is a marked ruler number.*

Proof. Exercise 9.1. ■

The second of our specific problems is that of constructing a cube root. Although more complicated solutions for constructing a cube root had been made earlier, Nicomedes, who was born about 270 BC, was apparently the first to use verging. In Figure 9.5, Nicomedes verges through C with respect to $\overleftrightarrow{AB}$ and $\overleftrightarrow{AD}$. Later solutions by Apollonius, Heron, Vieta, and Newton for the construction for a cube root are very much the same as that by Nicomedes when examined in detail.

Theorem 9.5 (Nicomedes). *Let* $\triangle ACB$ *have sides of lengths* $1,1,k/4$, *where* $AB = k/4$ *with* $0 < k < 8$. *Let* B *be the midpoint of* C *and* D. *Let* $\overrightarrow{CR}$ *intersect* $\overrightarrow{DA}$ *at* R *and* $\overrightarrow{BA}$ *at* S *such that* $RS = 1$ *but* $R \neq B$. *Then* $AS = \sqrt[3]{k}$.

Proof. Let the parallel to $\overleftrightarrow{AB}$ that passes through C intersect $\overleftrightarrow{DA}$ at E. So $\triangle ABD \sim \triangle ECD$. Then, since B bisects $\overline{DC}$, we have $CE = 2BA = k/2$. Also, since $\triangle ECR \sim \triangle ASR$, we have $(k/2)/CR = AS/1$. With $x = AS$, then $CR = k/(2x)$. With M the midpoint of A and B, by two applications of the Pythagorean Theorem, we now have

$$\begin{aligned}[1 + k/(2x)]^2 &= CS^2 = CM^2 + MS^2 = [CB^2 - BM^2] + MS^2 \\ &= [1^2 - (k/8)^2] + [x + (k/8)]^2.\end{aligned}$$

This reduces to the quartic equation $4x^4 + kx^3 - 4kx - k^2 = 0$. Fortunately, this quartic easily factors as $(4x+k)(x^3-k) = 0$. Since $4x+k > 0$, then we must have $x^3 - k = 0$. Therefore, x is the real cube root of k, as desired. ■

Corollary 9.6 (Cube Root Theorem). *If x is a marked ruler number, then $\sqrt[3]{x}$ is a marked ruler number.*

Proof. Exercise 9.2. ∎

By the Trisection Theorem and the Cube Root Theorem, we see that $\mathbb{V}$ is a real euclidean field such that $\sqrt[3]{x}$ is in $\mathbb{V}$ whenever x is in $\mathbb{V}$ and such that $\cos(x/3)$ is in $\mathbb{V}$ whenever $\cos x$ is in $\mathbb{V}$.

Definition 9.7. A field F is ***closed under cube root*** if x in F implies $\sqrt[3]{x}$ is in F. A field F is ***closed under trisection*** if $\cos x$ in F implies $\cos(x/3)$ is in F. A euclidean field is ***vietean*** if the field is closed under cube root and closed under trisection.

Since $\mathbb{V}$ is a euclidean field that is closed under cube root and closed under trisection, then $\mathbb{V}$ is vietean (pronounced vyā′-tē-en). That $\mathbb{V}$ is the smallest vietean field will follow after we have shown that all the real roots of a quadratic, cubic, or quartic with coefficients in a euclidean field closed under trisection and cube root are already in that field. Before proving this, we give a little of the bizarre background of polynomial equations with real coefficients.

The first degree equation $ax + b = 0$ with $a \neq 0$ has the solution

$$x = -\frac{b}{a}.$$

Today, we think it strange that there might ever have been some problem about accepting negative numbers. Without negative numbers, the general linear equation $ax + b = 0$ does not even make sense. As late as the middle of the eighteenth century, it was not clear to some mathematicians that it made sense to multiply two negative numbers. Descartes would at first not recognize what we call the cartesian plane; negative coordinates would be a confusing surprise. (How many of us feel unhappy outside the first quadrant of the cartesian plane? How many of us would respond to "Is $-x$ positive or negative?" without hesitation and yet without further explanation would identify "Is x positive or negative?" as a nonsense statement?)

The Babylonians devised the method of solving a quadratic equation by completing the square, which in modern notation gives the quadratic formula as the solution to the quadratic equation. The quadratic equation $ax^2 + bx + c = 0$ with $a \neq 0$ has the solution

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

Today, we probably have no problem with $b^2 - 4ac < 0$. In high school, we learned about complex numbers so that every quadratic would have roots. You may be surprised to learn that, unlike your own personal history, negative numbers and complex numbers were generally accepted at about the same time in mathematical history. The roots of the quadratic equation are expressed in terms of the coefficients, using the four arithmetic operations and radicals. In particular, the real roots are expressed in terms of the four arithmetic operations and real square roots.

As the Greek Menaechmus had given a geometric solution for duplicating the cube based on intersecting conics, the mathematician and astronomer Omar Khayyam (1050–1122), who is better known in the West as the Persian poet who wrote the *Rubaiyat*, gave general geometric solutions to cubics. For example, substituting the equation $x^2 = ry$ of a parabola into the cubic equation $x^3 + cx^2 + dx + e = 0$ gives $rxy + dx + cry + e = 0$, which is the equation of a hyperbola. The positive roots of the cubic are determined by the intersections of the conics. Omar Khayyam mistakenly believed that algebraic solutions for the general cubic were impossible. He also held the idea that geometric solutions for higher degree equations were impossible since space has but three dimensions. The publication of methods for solving cubics and quartics in *Ars Magna* in 1545 by Girolamo Cardano (Jerome Cardan, 1501–1576) is sometimes taken as the beginning of modern mathematics. The solution to the cubic followed the solution to the quadratic by nearly four thousand years. The story behind this unexpected event may be cloudy because of the unreliability of the cast of characters, but we will see that the story certainly does not lack color.

Cardano was one of the most fascinating of Renaissance men and without any doubt one of the most extraordinary characters in the history of mathematics, as he himself relates in his autobiography *Book of My Life*, which is a Dover reprint. His excessive gambling provided the background for his mathematical book *Book of Games of Chance*, which incorporates such practical advice as how to cheat to win. He was a successful physician, so famous that he was called to far away Scotland to diagnose the Archbishop of St. Andrews. Cardano, a bastard son of a jurist, is said to have disciplined his own wayward son by cutting off the son's ears. Cardano was imprisoned for casting a horoscope of Jesus and yet later hired by the Pope as an astrologer. The resourceful Cardano had learned that Tartaglia (the Stammerer, Niccolo Fontana, 1499–1557) had demonstrated in a public mathematics contest that he knew how to solve cubics. Tartaglia stammered from a saber cut in the face he had received as a child from a French soldier. Tartaglia revealed his method of solving cubics to Cardano in 1539, only after obtaining a pledge from Cardano to keep the method secret. In those days mathematical accomplishments were generally hoarded like military secrets and not rushed into publication. As we know, Cardano published the result in 1545. There are mitigating circumstances, however. Cardano and his student Ludovico Ferrari (1522–1565) claimed they learned in 1542 that Tartaglia's method was the same as that invented by Scipione dal Ferro (1465–1526) about 1500. Also, it is often overlooked in telling the story that Cardano did give credit to Ferro and to Tartaglia in the *Ars Magna*. Tartaglia and Ferrari took to quarreling about who first solved the cubic. Tartaglia is known to have plagiarized two other authors. Ferrari, by the way, is reported to have been poisoned by his sister. Cardano died by his own hand to fulfill his earlier astrological prediction of the date of his death. In any case, such claims and counterclaims were made in the dispute that mathematical historians are still trying to sort out the truth in order to determine who should get the credit for first solving the cubic.

The solution of the general quartic by Ferrari quickly followed his learning the solution of the general cubic. This too was published in 1545 in Cardano's *Ars Magna.*

The promised theorem that x is in $\mathbb{V}$ whenever x is any real solution of any quartic or cubic polynomial equation with coefficients in $\mathbb{V}$ is a corollary of the next theorem. The most interesting way to show that the roots can be found is to find them. The theorem is proved by providing an algorithm for finding roots of cubic and quartic polynomial equations. Since $\mathbb{R}$ is certainly a vietean field, then the proof of the next theorem also gives rules for solving cubics and quartics with any real (or complex) coefficients. The proof is longer than necessary but more interesting for the digressions.

Theorem 9.8. *Let V be a vietean field. If x is a real number such that*

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

where a,b,c,d,e are in V with $e \neq 0$, then x is in V.

Proof. If both a and b are 0, then the solutions follow from the quadratic formula and that field V is euclidean. We first suppose $a = 0$ but $b \neq 0$. Then, without loss of generality, we may suppose $b = 1$. Setting $x = y - c/3$ in our equation $x^3 + cx^2 + dx + e = 0$, we obtain the so-called depressed cubic

$$y^3 + py + q = 0,$$

where $p = d - c^2/3$ and $q = 2c^3/27 - dc/3 + e$. Then, x is in V iff y is in V. We can solve the original cubic in x iff we can solve the new cubic in y. Both cubics have coefficients in V; the equation $x = y - c/3$ tells us how to get from one solution to the other.

Our first solution will follow Cardano. This solution uses the identity

$$(A + B)^3 - 3AB(A + B) - (A^3 + B^3) = 0.$$

We set $-3AB = p$ and $-(A^3 + B^3) = q$. If we can find A and B that satisfy these two equations simultaneously , then $y = A + B$ will solve our problem because we will have $0 = y^3 - 3ABy - (A^3 + B^3) = y^3 + py + q$. The two equations $-3AB = p$ and $-(A^3 + B^3) = q$ give the identity

$$(z - A^3)(z - B^3) = z^2 - (A^3 + B^3)z + (AB)^3 = z^2 + qz - p^3/27$$

for all z. So the quadratic equations $(z - A^3)(z - B^3) = 0$ and $z^2 + qz - p^3/27 = 0$ have the same solutions. Equation $(z-A^3)(z-B^3) = 0$ has roots A^3 and B^3; equation $z^2+qz-p^3/27 = 0$ has roots $-q/2 \pm \sqrt{(q/2)^2 + (p/3)^3}$. Therefore, we can take $A^3 = -q/2 + \sqrt{(q/2)^2 + (p/3)^3}$ and $B^3 = -q/2 + \sqrt{(q/2)^2 + (p/3)^3}$. After all these tricks, it is now easy to find A and B. Let

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

We must check that

$$y_0 = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

is a solution to the cubic $y^3+py+q = 0$. In fact, if you know about complex cube roots, you can check that the three roots of the cubic in y are given by the three so-called *Cardano Formulas*:

$$A + B, \qquad \omega A + \omega^2 B, \qquad \omega^2 A + \omega B,$$

where $\omega = (-1+i\sqrt{3})/2$. (Since $\omega^3 = 1$, the three complex roots of 1 are 1, ω, and ω^2.) Since V is closed under cube root, if y_0 is in $\mathbb{R}$ then y_0 is in V. Other real solutions for y are also in V; these solutions are found by solving the quadratic equation with coefficients in V determined by dividing the cubic by the term $y - y_0$. (See Exercise 9.3.)

Our second method of solving the cubic is due to Vieta and is very elegant. We can suppose we have the equation $y^3 + py + q = 0$ with p and q in V. The ingenious substitution $y = p/(3z) - z$ gives the equation $z^6 - qz^3 - (p/3)^3 = 0$. Since this last equation is a quadratic in z^3 with coefficients in V, we can find in turn z^3, z, and then y. Any solution for z in $\mathbb{R}$ determines a solution for y in V. Exercise 9.13 asks you to derive the Cardano Formulas from the solution determined by Vieta's method.

There is one problem, however. Suppose $(q/2)^2 + (p/3)^3$ is negative. This has traditionally been called the *casus irreducibilis* or *irreducible case*, although, as we will see, the cubic has three real roots in this case. Even the Cardano Formula $A + B$ involves complex numbers in the irreducible case. Cardano, who would not admit a negative number as a "real" root of an equation, was perplexed about the appearance of these little understood complex numbers in his formulas. (Complex numbers are still called *imaginary numbers* and at that time negative numbers were called *fictitious numbers.*) Cardano said of the formula in this case that it is "as subtle as it is useless." The Cardano Formulas are correct if we use complex numbers but it is true that they are somewhat useless for computation using only real numbers. By using trigonometric identities and a clever substitution, Vieta showed the way to avoid imaginary numbers in the irreducible case.

Now, $(q/2)^2 + (p/3)^3 < 0$ iff $27q^2 < -4p^3$, in which case $p < 0$ and $|(-q/2)\sqrt{-27/p^3}| \leq 1$. So, in the irreducible case, $p < 0$ and there is an r such that $\cos(3r)^\circ = (-q/2)\sqrt{-27/p^3}$. Vieta's clever substitution

$$y = t\sqrt{-p/3}$$

in the equation $y^3 + py + q = 0$ then gives the equation

$$t^3 - 3t - 2\cos(3r)^\circ = 0,$$

which has roots

$$2\cos r^\circ, \quad 2\cos(r + 120)^\circ, \quad 2\cos(r + 120)^\circ,$$

by Lemma 2.20. So if $(q/2)^2 + (p/3)^3 < 0$ and $\cos(3r)^\circ = (-q/2)\sqrt{-27/p^3}$, then $y^3 + py + q = 0$ has the roots

$$\sqrt{-4p/3}\,\cos(r + 120k)^\circ, \quad k = 0, 1, 2.$$

All these roots are real. Since V is closed under trisection, then the roots are in V.

A summary of the cubic case may be in order. The depressed cubic $y^3 + py + q = 0$ with real coefficients p and q has either only one real root, which is given by the Cardano Formula $A + B$, or else three real roots, which are expressed above in terms of the cosine function. If p and q are in V, then the real roots are in V. Some of the unsatisfactory aspects of these solutions are illustrated in Exercises 9.14–9.16. It can be shown, although we have not done so, that in the general irreducible case the three real roots cannot be expressed in terms of the coefficients and real radicals. In complex numbers, however, all three roots are given by the Cardano Formulas.

It does seem paradoxical that the limitation of not being able to express the roots of an arbitrary depressed real cubic in terms of real radicals applies to the case where all three roots are real. For example, it can be shown that such is the case for $x^3 - 2px + p = 0$ where p is any prime. However, for our purposes, the most prominent example is the equation $x^3 - 3x - 1 = 0$, which has three real roots $2\cos 20^\circ$, $2\cos 140^\circ$, $2\cos 260^\circ$. Although it is impossible to write $2\cos 20^\circ$ in terms of real radicals, we can use the Cardano Formulas to obtain $2\cos 20^\circ$ in terms of complex radicals:

$$2\cos 20^\circ = \sqrt[3]{\frac{1 + \sqrt{-3}}{2}} + \sqrt[3]{\frac{1 - \sqrt{-3}}{2}}.$$

We now turn to the quartic $ax^4 + bx^3 + cx^2 + dx + e = 0$ with $a \neq 0$. Without loss of generality, we may suppose $a = 1$. Further, by the substitution $x = y - b/4$, we have the depressed quartic

$$y^4 + py^2 + qy + r = 0.$$

If $q = 0$, then we have a quadratic in y^2 and the solutions are quickly obtained. Otherwise, following the ideas of Vieta, we rewrite this equation as $y^4 = -py^2 - qy - r$ and then add $zy^2 + z^2/4$ to both sides to obtain

$$(y^2 + z/2)^2 = (z - p)y^2 - qy + (z^2/4 - r).$$

The left-hand side of this equation is a square. If z is a number such that the right-hand side is the square of a term of the form $gy+h$, then the solutions for y will follow easily from the equations $y^2 + z/2 = \pm(gy + h)$. Hence, we want z to be such that the right-hand side of the equation displayed above is a square, which is the case iff the roots of the quadratic

$$(z - p)y^2 + (-q)y + (z^2/4 - r) = 0$$

in y are equal. This happens iff $[-q]^2 - 4[z - p][z^2/4 - r] = 0$, which we rewrite as

$$z^3 - pz^2 - 4rz - (q^2 - 4pr) = 0.$$

The solution is now clear. It would be more accurate to say that *the method* for determining the solution is now clear. That is, we first solve this cubic for a root z with $z \geq p$ by the methods from above and then solve quadratic equations for y. Generally, in order to solve a quartic, we must first solve a cubic. Nobody said it was going to be easy to solve the general quartic equation. On the other hand, it is truly amazing that any quartic can be solved just because any cubic can be solved. Since V is closed under trisection and under cube root, then all real solutions of the quartic are in V.

Another version of the same idea for solving a quartic runs as follows. Starting with $x^4 + bx^3 + cx^2 + dx + e = 0$, we have $x^4 + bx^3 = -cx^2 - dx - e$. Completing the square on the left, we have $(x^2 + bx/2)^2 = (b^2/4 - c)x^2 - dx - e$. Adding $(x^2 + bx/2)y + y^2/4$ to each side, we get

$$(x^2 + bx/2 + y/2)^2 = (b^2/4 - c + y)x^2 + (by/2 - d)x + (y^2/4 - e).$$

Again, we want a number y such that the right-hand side is a square. You can check that this requires that y be a root of

$$y^3 - cy^2 + (bd - 4e)y + (4ce - b^2e - d^2) = 0.$$

This method is due to Ferrari. ∎

Corollary 9.9. *If x is a real root of a polynomial equation of degree 4 or less and the coefficients of the equation are in $\mathbb{V}$, then x is in $\mathbb{V}$.*

We saw above that being able to solve linear, quadratic, and cubic polynomial equations automatically means being able to solve quartic polynomial equations, without the requirement of any new operations. The construction problems that can be solved with the marked ruler are precisely those that when transcribed into algebra have solutions in $\mathbb{V}$ that arise from linear, quadratic, cubic, or quartic polynomial equations with coefficients in $\mathbb{V}$. A construction problem that was solved in antiquity in terms of the intersection of two conic sections translates today into an algebraic problem that requires finding solutions to a polynomial equation of fourth degree or less. Thus, the conic solutions of the Greeks can be achieved as marked ruler constructions.

Since the coefficients of $x^3 + x^2 - 2x - 1 = 0$ are certainly in $\mathbb{V}$ and since $2\cos(360/7)^\circ$ is a root of this cubic by Theorem 2.23, then the regular heptagon is a marked ruler construction. We state this result formally in our next corollary. See Exercises 9.11 and 9.12 for constructions, especially Plemelj's elegant construction of 1912 which is given in The Back of the Book for 9.12. Of course, the reguler enneagon is also a marked ruler construction by Theorem 9.3.

Corollary 9.10. *The numbers* $\cos(360/7)^\circ$ *and* $\cos(360/9)^\circ$ *are in* $\mathbb{V}$ *but not in* $\mathbb{E}$.

The history of the first solution to the construction of the regular heptagon is complicated and reminds us of the priority arguments almost six hundred year later over the solution of the cubic equation. The heptagon construction was published by three Arab geometers almost simultaneously. The date is probably the year 969. The story, which mainly concerns a bitter dispute between two young geometers who were both engaged in plagiarism, may be as follows. Abū'l-Jūd initially gave a partially erroneous solution to the construction. Al-Sijzī discovered this error and then requested Al-^{c}Alā to solve a problem that would complete Abū'l-Jūd's solution. Without knowing the purpose of the problem that had been given to him, Al-^{c}Alā unknowingly provided the missing link. Ironically, in his response to astrologer and geometer Al-Sijzī, Al-^{c}Alā gave his opinion that a conic construction of the regular heptagon is impossible. Abū'l-Jūd and Al-Sijzī then both published solutions without mention of Al-^{c}Alā and accused the other of plagiarism. Very shortly thereafter, Al-Kūhī, who previously had been a juggler of glass bottles in the market place of Baghdad, was the third to give a conic solution along with a detailed analysis of Archimedes' treatment of the problem. These great Arab achievements occurred some twelve hundred years after the time of Archimedes. Without comment, we note that Al-Sijzī lamented that where he lived the people considered it lawful to kill mathematicians.

The classical Arabic literature contains at least a dozen more conic constructions for the construction of the regular heptagon. Several of these are due to the great Arab scientist and mathematician Ibn al-Haytham (*circa* 965–1040), who is known to the West as Alhazen, from the Latinized form of his first name al-Hasan.

The construction of the regular heptagon did not receive much attention in Greek geometry. Other than an approximation by Heron, the heptagon

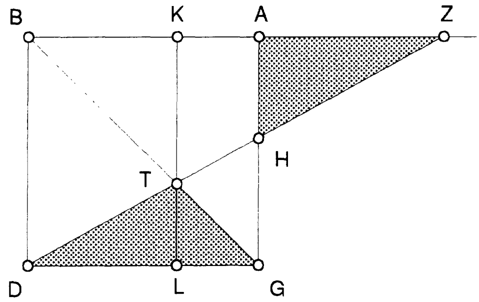


FIGURE 9.6.

is hardly mentioned. However, there is one remarkable "construction" attributed to Archimedes by the Arabs. There exists a "corrected" and edited manuscript from about 1740 of an earlier translation by Thābit ibn Quarra (830–901) of *Book of Construction of the Circle Divided into Seven Equal Parts by Archimedes.* This book contains eighteen propositions, with the first sixteen concerning right triangles and trigonometry. The title apparently comes from the last two propositions. The construction of the regular heptagon in Proposition 18 is based on Proposition 17, which the Arab geometers called *the Lemma of Archimedes.* It is the questionable construction provided in Proposition 17, which is unlike anything else in Greek mathematics, that stimulated the Arab mathematicians to devise conic constructions of the regular heptagon, a task never accomplished by the Greeks.

Given $\overline{AB}$, the Lemma of Archimedes asks for the construction of points K and Z with B–K–A and B–A–Z such that $(BA)(BK) = ZA^2$ and $(KZ)(KA) = KB^2$. Suppose $\square ABDG$ is a square. See Figure 9.6. Let T be the point on diagonal $\overline{BG}$ such that $\overrightarrow{DT}$ intersects $\overline{AG}$ at H, intersects $\overrightarrow{BA}$ at Z, and such that $DTG = AHZ$. (Although ingenious, this "construction" presents more problems than it solves. Point T certainly exists but we are not told how to find T such that triangles $\triangle DTZ$ and $\triangle AHZ$ have the same area. Note that if we take $AB = 1$ and $AZ = x$, then $BK = x^2$ with $x^3 + 2x^2 - x - 1 = 0$. So, x is in $\mathbb{V}$ but not in $\mathbb{E}$.) Let the parallel to $\overline{AG}$ through T intersect $\overline{BA}$ at K and intersect $\overline{DG}$ at L. We now prove that K and Z are as desired. Since $(BA)(LT) = (DG)(LT) = 2\,DTG = 2\,AHZ = (AH)(ZA)$, then $BA/ZA = AH/LT$; since $\triangle AHZ \sim \triangle LTD$, then $AH/LT = ZA/DL = ZA/BK$. Thus, $(BA)(BK) = ZA^2$. Since $KA = LG = TL$ and $LD = KB = TK$, then $KA/KB = TL/TK$; since $\triangle TLD \sim \triangle TKZ$, then $TL/TK = LD/KZ = KB/KZ$. Thus, $(KZ)(KA) = KB^2$, as desired.

Proposition 18 is something of a muddle and we follow the analysis of Al-Kūhī to construct an angle of $(360/7)^\circ$ by constructing a triangle whose angle measures are in the ratio $1 : 2 : 4$. Assuming the Lemma of Archimedes and the notation above, we claim that circles K_{BK} and A_{ZA} intersect. See

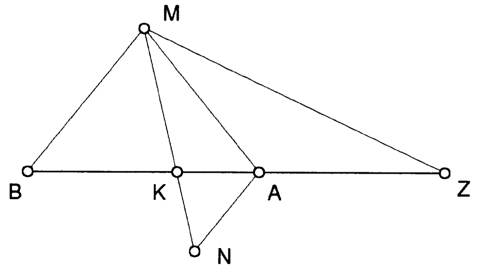


FIGURE 9.7.

Figure 9.7. First, from the definition of T, we have $BK+AZ > BK > AK$ and $BK+KA = BA > AZ$. Also, since $(BA)(BK) = ZA^2$ and $BA > BK$, then $ZA > BK$ and so $KA + ZA > ZA > BK$. It follows that K_B and A_Z intersect, say at point M. Let $t = m\angle BZM$. Then $m\angle AMZ = t$ and $m\angle MAK = 2t$. Next, since $(KZ)(KA) = KB^2 = KM^2$, then $KZ/KM = KM/KA$. So $\triangle KZM \sim \triangle KMA$ and $\angle KZM \cong \angle KMA$. Hence, $m\angle KMA = t$. We will be done after we show $m\angle KMB = 2t$, as then $\triangle BMZ$ will have angles of measure t, $2t$, and $4t$.

Let N be such that M–K–N and $KN = KA$. Since $\triangle BKM \sim \triangle AKN$ with both triangles isosceles, the angles $\angle KBM$, $\angle KMB$, $\angle KAN$, and $\angle KNA$ are congruent to each other. We next show that each of these angles has measure $2t$. By the Lemma of Archimedes, we know $MA^2 = ZA^2 = (BA)(BK) = (MN)(MK)$. So $MA/MN = MK/MA$ and then $\triangle MAN \sim \triangle MKA$ and $\angle MNA \cong \angle MAK$. Hence, $m\angle MNA = 2t = m\angle KAN$, as desired. Therefore, we now have $m\angle BZM = t$, $m\angle ZBM = 2t$, and $m\angle BMZ = 4t$. Hence, $t = 180/7$ and we are done. We note that $\overline{BM}$ is the side of a regular heptagon inscribed in the circle through the points B,M,Z. Al-Kūhī goes on to show that the points given by the Lemma of Archimedes can be obtained as conic intersections.

Although Alhazen (Ibn al-Haytham) wrote a commentary on Euclid's *Elements*, two books on squaring lunes, and one entitled *Quadrature of the Circle*, mathematicians know him for one particular problem from his best-known work, the *Treasury of Optics.* This problem, which we will examine below, is known as Alhazen's Problem and concerns the path of light reflected by a circular mirror. Perhaps the most interesting story about Alhazen is that this already famous mathematician was called to Cairo from his home in Basra, Iraq, in order to carry out his plan for controlling the flow of the Nile by building a dam across the river near Aswan. However, once in Egypt, Alhazen realized that his plan was not feasible and admitted his failure to the capricious and murderous caliph. Fearing the insecurity of his position, Alhazen pretended to be mentally deranged and was confined to his house until the caliph's death. After the caliph died, Alhazen resumed his writing and teaching, surviving in Cairo for twenty more years.

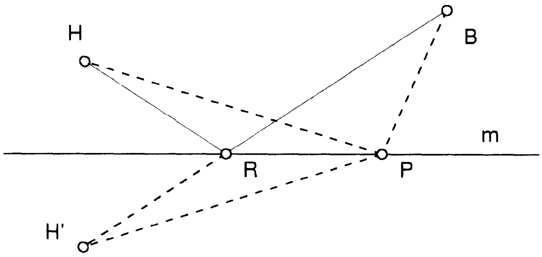


FIGURE 9.8.

Given points H and B on the same side of line m, the construction of the shortest path from H to m to B is a ruler and compass construction. See Figure 9.8. As Heron of Alexandria argued in his *Theory of Mirrors* about the year 100, if R is the point on m such that $HR + RB$ is minimal then R is the intersection of m and the line connecting B and H', where H' is the image of H under the reflection in m. To see this, suppose P is any point on m. Since m is the perpendicular bisector of $\overline{HH'}$, then $HP = H'P$. So $HP + PB = H'P + PB$. Since we now want $H'P + PB$ to be minimal, then we need H', P, and B to be collinear. The path has minimum length when $P = R$. Considering m to be the surface of a mirror, our path from H to R to B is that of light traveling from H to B by reflection in m at R. (Note the two very different meanings of "reflection in m.")

Given points H and B outside circle m (or both inside m), *Alhazen's Problem* is to construct, when possible, the shortest path from H to m to B. We want the path that light would take traveling from H to B by reflecting in circular mirror m. See Figure 9.9. We will see that this is not a ruler and compass construction. Alhazen was the first to show that this problem can be solved with conics. A thousand years ago, Alhazen used a long sequence of complicated lemmas for his demonstration. This was a considerable achievement; we turn to coordinate geometry. Without loss

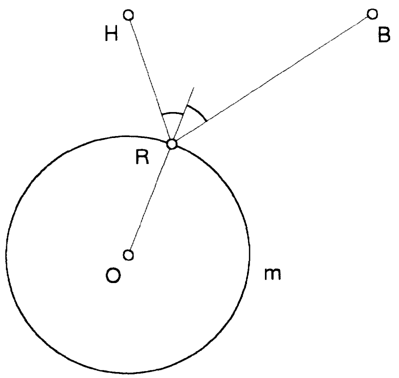


FIGURE 9.9.

of generality, we may suppose m is the unit circle, $H = (0, a)$, $B = (b, c)$ with $b \neq 0$, and R is the desired point of reflection. Let $R = (x, y)$ and $O = (0, 0)$. So $x^2 + y^2 = 1$. We recall from analytic geometry (or check any calculus text) that if θ is the measure of an angle from a line with slope m_1 to a line with slope m_2 then $\tan\theta^\circ = (m_2 - m_1)/(1 + m_2 m_1)$. In our case, we want an angle from $\overleftrightarrow{RH}$ to $\overleftrightarrow{RO}$ to be congruent to an angle from $\overleftrightarrow{RO}$ to $\overleftrightarrow{RB}$. Since the lines $\overleftrightarrow{RH}$, $\overleftrightarrow{RO}$, $\overleftrightarrow{RB}$ have slopes $(y - a)/x$, y/x, $(y - c)/(x - b)$, respectively, we have the formula

$$\frac{\frac{y-a}{x} - \frac{y}{x}}{1 + \frac{(y-a)y}{x^2}} = \frac{\frac{y}{x} - \frac{y-c}{x-b}}{1 + \frac{y(y-c)}{x(x-b)}},$$

which reduces to the equation $y[2acx + b] = ab + (a + c)x - 2abx^2$. Squaring both side of this equation and replacing y^2 by $1 - x^2$, we get a quartic polynomial equation in x with coefficients in terms of a, b, c. Therefore, by Corollary 9.9, we know that the construction is a marked ruler construction.

We want to show that Alhazen's Problem is in general not a ruler and compass construction. Quartics are usually not easy to factor. However, we can arrange to have one of the extraneous roots of the quartic be $+1$, which is picked up because the angles satisfying the formula differ by 180 in measure. We arrange this by taking $a = b = c = 2$. In this case, after discarding the factor $x - 1$ from the quartic, we arrive at the cubic equation $32x^3 + 24x^2 - 3x - 3 = 0$. This cubic has three real roots, one of which is positive, but no root in $\mathbb{Q}$ and, therefore, no root in $\mathbb{E}$. In general, Alhazen's Problem is not a ruler and compass construction.

If we know that light traveling from one focus of an ellipse and reflected in the ellipse travels directly to the other focus, the solution to Alhazen's Problem determines the ellipse that has foci H and B and is tangent to circle m. The point of tangency is R, the point of reflection of light traveling from H to B. Although we now know that it is possible to construct R with a marked ruler by a long sequence of constructions that essentially solve the quartic equation, we admit that we have not produced a nice marked ruler construction for Alhazen's Problem. In this, we are no further ahead than we were a thousand years ago.

The solution to cubics and quartics by radicals in complex numbers gave hope that solutions could be found for the general quintic. The last chapter in the history of the search for solutions to the polynomial equations is another illustration of two mathematicians independently solving a problem at roughly the same time. This is another case where the problem is solved by showing that a solution is impossible. For the general polynomial equation of degree 5 or higher, there is no formula or algorithm for expressing a root in terms of the coefficients and algebraic operations (the four arithmetic operations and extracting k^{th} roots for complex numbers). For example, if p and q are primes with $q \geq 5$ and a is an integer such that $a \geq 2$, then it can be shown that the polynomial equation $x^q - apx - p = 0$

cannot be solved in complex radicals. The simplest special case is the quintic equation $x^5 - 4x - 2 = 0$.

In 1799, Paolo Ruffini (1765–1822) gave a proof that virtually established the unsolvability of the general quintic by radicals. Ruffini's definitive work of 1813 on the subject is essentially the same as Wantzel's simplification of Abel's proof of 1822. The Norwegian Niels Henrik Abel (1802–1829) published this proof in 1824. Remediable defects mar the proofs of both Ruffini and Abel. Although Ruffini's work was overlooked for some time, the impossibility of solving general polynomial equations of degree higher than 4 is now referred to as the Abel–Ruffini theorem. This brought to an end the search for formulas to solve general polynomial equations, a search begun by the Babylonians nearly four thousand years earlier.

In the second volume of the *American Mathematical Society Bulletin* (December 1895), Yale Professor James Pierpont (1866–1938) points out that Gauss did not completely prove what we in Chapter 2 have called the Gauss–Wantzel Theorem. He then sets about to fill the hole left by Gauss, without any mention whatsoever of Wantzel. Therefore, the paper would be of no significance if Pierpont had not gone on to make an important observation for which we need the following definition. A *pierpont prime* is a prime of the form $2^a3^b + 1$. Pierpont primes are nowhere near as rare as fermat primes, but more about that below. Pierpont observed that the Greeks frequently used conic sections to give geometric constructions and asked the question, What regular polygons are constructible if the use of rational conic sections is allowed? We have seen, by a theorem of Vieta, that the use of rational conic sections is equivalent to being able to trisect angles and take cube roots. Thus we would rephrase Pierpont's question as, What regular polygons can be constructed with the marked ruler? By adjusting the proof of the Gauss–Wantzel Theorem slightly, Pierpont reaches the following answer.

Pierpont's Theorem. *A regular n-gon is constructible with the marked ruler iff $n > 2$ and $n = 2^a3^bp_1p_2\dots p_k$ where a,b,k are nonnegative integers and the p_i are distinct pierpont primes.*

We give the following table, which ends Pierpont's article and which should be self-explanatory. We are looking at n-gons with $n \leq 100$.

Greeks: 2, 3, 4, 5, 6, 8, 10, 12, 15, 16, 20, 24, 30, 32, 40, 48, 60, 64, 80, 96.

Gauss: 17, 34, 51, 68, 85.

Conics: 7, 9, 13, 14, 18, 19, 21, 26, 27, 28, 35, 36, 37, 38, 39, 42, 45, 52, 54, 56, 57, 63, 65, 70, 72, 73, 74, 76, 78, 81, 84, 90, 91, 95, 97.

With the obvious, reasonable definitions, it follows from the Steiner–Poncelet Theorem that the tomahawk of Chapter 1 is equivalent to ruler, compass, and angle trisector. The field $\mathbb{T}$ of tomahawk numbers is then the smallest euclidean field that is closed under trisection. In other words, $\mathbb{T}$ is the intersection of all fields that are closed under trisection. We have seen above that angle trisection corresponds to solving Cardano's irreducible case for the cubic equation. The geometric tool called the tomahawk corresponds to the euclidean field $\mathbb{T}$ where cubics having coefficients in $\mathbb{T}$ and having three real roots necessarily have all three roots in $\mathbb{T}$. American mathematician Andrew M. Gleason has recently sharpened Pierpont's Theorem, showing that the cube root construction is not needed for constructing the regular polygons. We state this result in the following form.

Gleason's Theorem. *A regular polygon is constructible with the marked ruler iff the regular polygon is constructible with the tomahawk.*

So the regular heptagon (7-gon) and the regular triskaidecagon (13-gon) are tomahawk constructions. The regular hendecagon (11-gon) is not a tomahawk construction, however, since 11 is not a pierpont prime. A fermat prime is a pierpont prime. Conjecturing that there are infinitely many pierpont primes, Gleason gives the 41 pierpont primes that are less than one million: 2, 3, 5, 7, 13, 17, 19, 37, 73, 97, 109, 163, 193, 257, 433, 487, 577, 769, 1153, 1297, 1459, 2593, 2917, 3457, 3889, 10369, 12289, 17497, 18433, 39367, 52489, 65537, 139969, 147457, 209953, 331777, 472393, 629857, 746497, 839809, 995329.

Gleason goes on to remark that a generalization of a corollary from Gauss's *Disquisitiones Arithmeticæ* yields the following theorem, where $\phi(n)$ denotes the number of positive integers less than n that are relatively prime to n: A regular n-gon can be constructed if, in addition to ruler and compass, tools are available to p-sect any given angle for each prime p that divides $\phi(n)$. For example, since $\phi(11) = (2)(5)$, then ruler, compass, and quinquesector are sufficient tools to construct a regular hendecagon.

The field $\mathbb{G}$ of *glotin numbers* corresponds to the geometric construction tool consisting of the tomahawk together with an arbitrary number of attachable T's. This extension of the tomahawk by Glotin is described in Chapter 1; see Figure 1.11c. The tomahawk with attachable T's allows n-section of any given angle for any positive integer n. All regular n-gons are constructible with this tool. We may as well mention the field $\mathbb{H}$, the *field of Hippias*, associated with ruler, compass, and quadratrix. Recall from Chapter 2 that with the quadratrix all angle division problems are reduced to segment divisions. Here, for any positive number x, if we can x-sect a segment then we can x-sect a given angle. Finally, if we augment the quadratrix with its limit point, the point H in Figure 2.7, then the tools are now associated with the field that is denoted by $\mathbb{H}(\pi)$. As we have seen in Chapter 2, with these tools we can now square the circle. We mention only that the extension $\mathbb{H}(\pi)$ of the field $\mathbb{H}$ is not a quadratic extension but is of such a different nature from anything else we have seen so far mentioned that the extension is called *transcendental.*

A tomahawk point is certainly a marked ruler point. Obviously, $\mathbb{E} \subset \mathbb{T} \subseteq \mathbb{V}$. However, $\mathbb{T} \neq \mathbb{V}$ as $\sqrt[3]{2} \notin \mathbb{T}$. (Note that the equation $x^3 - 2 = 0$ does not have three real roots.) The Delian Problem is unsolvable with the tomahawk.

Whether realized as a tool based on Nicomedes' cube root construction of Figure 9.5 or based on the cissoid of Diocles, we can suppose that we have an ideal "cube rootor," the geometric tool that allows the construction of a segment of length the cube root of the length of a given segment. The field of numbers associated with ruler, compass, and cube rootor is $\mathbb{D}$, the smallest euclidean field that is closed under cube root. Since $\cos 20°$ is not in $\mathbb{D}$, then both $\mathbb{T}$ and $\mathbb{D}$ are properly contained between $\mathbb{E}$ and $\mathbb{V}$. Of course, $\mathbb{V}$ is the smallest field that contains both $\mathbb{D}$ and $\mathbb{T}$.

The principal topic of this book is the association of fields and geometric construction tools. The fields allow us to characterize those constructions that are possible with the given tools. We can easily imagine an abundance of combinations of geometric construction tools such that each of these combinations give rise to a field of numbers. However, we need not resort to bizarre tools such as the angle quinquesector and the cube rootor to find a topic that is worthy of further study. Surprisingly, there seems to be little published on marked ruler and compass constructions. Using the compass along with the marked ruler opens the possibility of verging with respect to a line and a circle as well as verging with respect to two circles. Archimedes' trisection of an angle uses verging with respect to a line and a circle, but this particular use is unnecessary since Pappus' angle trisection requires only verging with respect to two lines. Therefore, marked ruler and compass constructions are unknown territory. Surely $\mathbb{M}$, the field of marked ruler and compass numbers, is larger than $\mathbb{V}$, but what is the best algebraic description of $\mathbb{M}$? What new constructions are possible with these readily available tools? Is the hendecagon a marked ruler and compass construction?

Exercises

9.1. Prove Corollary 9.4, the Trisection Theorem.◊

9.2. Prove Corollary 9.6, the Cube Root Theorem.◊

9.3. If $x^3 + px + q = 0$ has a root r, then find the other two roots for the cubic.◊

9.4. Outline a marked ruler construction for a perpendicular to a given line.◊

9.5. Outline a marked ruler construction for bisecting a given segment.◊

9.6. Outline a marked ruler construction for a $60°$ angle.◊

9.7. Can a regular polygon with 27 sides be constructed with a marked ruler? Can an angle of $1°$ be constructed with a marked ruler?◊

9.8. Make a construction drawing for a $20°$ angle using a marked ruler and, to facilitate drawing perpendiculars, a right-angle ruler.◊

9.9. Find the smallest pierpont prime that is greater than one million.

9.10. Give a cubic equation whose positive root s is such that verging through $(-2/3, 1/3)$ with respect to the line with equation $2Y = -3X$, which contains R, and the X-axis, which contains S, we have $S = (s, 0)$.◊

9.11. Show the construction in Exercise 9.10 can be used in a marked ruler construction of a regular 7-gon.◊

9.12. Show $y^6 - 7y^4 + 14y^2 - 7 = 0$ has root s_7 in $\mathbb{V}$. (This cubic equation in y^2 leads to a proof of the tomahawk construction for a regular heptagon that is given in The Back of the Book.)◊

9.13. Show Vieta's method that uses the substitution $y = p/(3z) - z$ does give the Cardano Formulas and only these solutions.◊

9.14. Find x in $\mathbb{V}$ such that $x^3 + 3x + 14 = 0$ by using the Cardano Formulas.◊

9.15. Express the real root of $x^3 - 6x - 4 = 0$ as given by the Cardano Formulas. Solve the cubic using Vieta's method for the irreducible case. Express all roots in terms of real radicals.◊

9.16. Find all real roots of the equation $x^3 - 7x + 6 = 0$.◊

9.17. Find all real roots of the equation $x^3 + x^2 + x - 1 = 0$.◊

9.18. Find the length of the third side of the triangle having one side of length 5, one side of length 7, and an inscribed circle with radius of length 1. Show that this triangle can be constructed with the marked ruler but cannot be constructed with ruler and compass.◊

9.19. Recall the following definitions from calculus:

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2},$$

$r = \operatorname{arcsinh} k$ iff $\sinh r = k$, and $t = \operatorname{arccosh} k$ iff $\cosh t = k$. Prove the identities

$$4 \sinh^3 x + 3 \sinh x = \sinh 3x \quad \text{and} \quad 4 \cosh^2 x - 3 \cosh x = \cosh 3x.$$

9.20. Let $h = \sqrt{4|p|/3}$ and $k = (-q/2)\sqrt{27/|p|^3}$. Then prove:

Theorem A. If $p > 0$ and $y = h \sinh[(1/3)(\operatorname{arcsinh} k)]$, then $y^3 + py + q = 0$.

Theorem B. If $p < 0$, $k \geq 1$, and $y = h\cosh[(1/3)(\operatorname{arccosh} k)]$, then $y^3 + py + q = 0$.

Theorem C. If $p < 0$, $k \leq -1$, and $y = -h\cosh[(1/3)(\operatorname{arccosh}(-k))]$, then $y^3 + py + q = 0$.

Theorem D. If $p < 0$, $|k| \leq 1$, and $y = h\cos[(1/3)(\arccos k)]$, then $y^3 + py + q = 0$.

10
Paperfolding

> Conic sections became an intrinsic part of our culture when Kepler discovered that the planet Mars travels around the sun in an ellipse, the sun being at one focus.
>
> DAN PEDOE

Paperfolding, as a means of geometric construction and as opposed to origami, was introduced in 1893 by T. Sundara Row from India. In Sundara Row's *Geometric Exercises in Paper Folding* it is evident that all ruler and compass constructions are possible by paperfolding. However, Sundara Row's angle trisection is admittedly only an approximation and he mistakenly implies that constructing a cube root is impossible in general and, in particular, that the duplication of the cube cannot be accomplished by paperfolding. In his 1949 *A History of Geometric Methods*, J. L. Coolidge refers to Sundara Row but limits his own folding operations to those equivalent to the use of ruler alone. The first rigorous treatment of paperfolding is apparently by R. C. Yates in 1949 in his *Geometric Tools*. Yates postulated three operations: (i) place one point of the sheet upon another and thus create a crease; (ii) establish the crease through two given distinct points; and (iii) place a given point upon a given line so that the resulting crease passes through a second given point, when the given points and line are so situated that this may be accomplished. We will verify below that Yates's third postulate allows the construction of lines through a given point and tangent to the parabola with a given focus and given directrix. Yates proves:

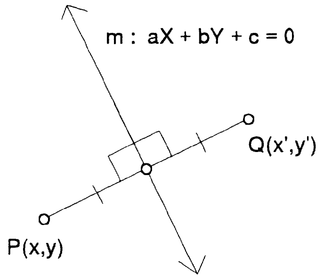


FIGURE 10.1.

Under the chosen postulates, all constructions of plane Euclidean geometry can be executed by means of creases.

We suppose our paper is transparent. Kitchen waxed-paper is ideal for paperfolding constructions. The crease of a fold is readily apparent upon unfolding the waxed-paper. Tracing paper is also good, but expensive. You might pause here to think of the common geometric constructions that can be accomplished by paperfolding.

The basic idea of paperfolding is that of a reflection. That is, after all, what a fold is all about. We have used the ideas of symmetry and reflection before, especially in developing the Mohr–Mascheroni theory of compass constructions. We are going to be a little more formal about things here and give a concise definition of a reflection while introducing some notation. In Figure 10.1, point Q is the image of P under the reflection in the line m. Although a young child can point to where Q should be, the standard formal definition takes a bit of understanding. This is mostly because the definition uses two conditionals. The definition has the same form as the definition of the absolute value function, which is usually the first definition encountered that uses two conditionals, where

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0, \end{cases}$$

for real number x.

Definition 10.1. Given line m, the ***reflection*** r_m is the mapping on the set of points in the plane such that for point P

$$r_m(P) = \begin{cases} P & \text{if } P \text{ is on } m, \\ Q & \text{if } P \text{ is off } m \text{ and } m \text{ is the perpendicular bisector of } \overline{PQ}. \end{cases}$$

The images of point P and line l under the reflection in line m are also denoted by P^m and l^m, respectively. That is,

$$P^m = r_m(P) \quad \text{and} \quad l^m = r_m(l) = \{r_m(P) \mid P \text{ is on } l\}.$$

Is it clear that $P^t = Q$ iff $P = Q^t$? If $P \neq Q$, then this important equivalence follows because the perpendicular bisector of $\overline{PQ}$ is the perpendicular bisector of $\overline{QP}$. While we are thinking about the definition of a reflection, we should compute the formulas for the coordinates for an image of a point under a reflection in the cartesian plane.

Theorem 10.2. *In the cartesian plane, the image of (x, y) under the reflection in the line with equation $aX + bY + c = 0$ is (x', y') where*

$$x' = x - \frac{2a(ax + by + c)}{a^2 + b^2} \quad \text{and} \quad y' = y - \frac{2b(ax + by + c)}{a^2 + b^2}.$$

Proof. The equations hold when (x, y) is on the line as then $(x', y') = (x, y)$, since (x, y) is on the line with equation $aX + bY + c = 0$ iff $ax + by + c = 0$. Now suppose that (x, y) is off the line. Then, by the definition of a reflection, the line m is the perpendicular bisector of the segment with endpoints (x, y) and (x', y'). So, (i) the midpoint of (x, y) and (x', y') is on the line m, which has equation $aX + bY + c = 0$; and (ii) the line through (x, y) and (x', y') is perpendicular to this line m. These two geometric properties are expressed, respectively, as the two algebraic equations

$$a[(x + x')/2] + b[(y + y')/2] + c = 0, \quad a(y' - y) = b(x' - x).$$

Rewriting these two equations as

$$\begin{cases} ax' + by' = -2c - ax - by, \\ bx' - ay' = bx - ay, \end{cases}$$

we see we have two linear equations in the two unknowns x' and y'. Solving these (Exercise 10.10), we obtain the formulas in the statement of the theorem. ∎

We will more easily understand what is going on later if we can answer the following question. What are the lines t such that the image of given point P under the reflection in t is on given line p? This question turns out to have a very nice answer, as we will see in the next theorem. Recall that a *parabola* is the locus of all points that are equidistant from a point P, called the *focus* of the parabola, and a line p, called the *directrix* of the parabola.

If at all possible, you should do some paperfolding as you read this chapter. Here, it is instructive to take a 12 inch square of waxed-paper, fold a line p near one edge of the square (or mark a straight-cut edge as line p), and mark a point P about 1.5 inches from p. Now, fold along a line such that in the folding P falls on p, unfold, and then repeat this folding process many times, folding about a different line each time. You should begin to see what looks like a parabola appearing; what you are doing is described as forming a parabola as the envelope of its set of tangents. One thing you

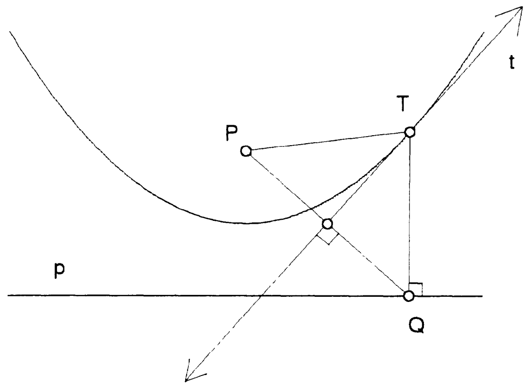


FIGURE 10.2.

will quickly learn is to take advantage of the mathematical equivalence of "P^t on p" and "P on p^t." Rather than have P fall on p in the folding above, you can have p fall on P. It amounts to the same thing mathematically, only the latter is easier to manage physically when p is an edge of the waxed-paper. After you have done this, you should continue experimenting with the waxed-paper to devise some paperfolding constructions of your own.

Theorem 10.3. *If point P is on line p, then P^t is on p iff t passes through P or t is perpendicular to p. If point P is off line p, then P^t is on p iff t is a tangent to the parabola with focus P and directrix p.*

Proof. The first part follows directly from the definition of a reflection.

For the second part, suppose P is off p and $P^t = Q$ with Q on p. By the definition of a reflection in line t, then t is the perpendicular bisector of $\overline{PQ}$. Since P is off p, then t and p are not perpendicular. The perpendicular to p at Q intersects t at some point T and $TP = TQ$. See Figure 10.2. So T is equidistant from the point P and the line p. In other words, T is on the parabola with focus P and directrix p. Assume S is a second point on t that is also on this parabola, and let F be the foot of the perpendicular from S to p. Then $F \neq Q$. Also, $SQ = SP$ since S is on t, the locus of all points equidistant from P and Q. Further $SP = SF$ since S is on the parabola, the locus of all points equidistant from P and p. Hence, $SQ = SF$ and we have the impossibility that $\triangle FSQ$ is an isosceles triangle with a right angle as a base angle. Assuming the existence of S has led to a contradiction. Thus, t contains exactly one point of the parabola; line t is a tangent of the parabola.

Conversely, suppose t is a tangent to the parabola with focus P and directrix p. Let t intersect the parabola at T, let Q be the foot of the perpendicular from T to p, and let n be the perpendicular bisector of $\overline{PQ}$. So Q is the image of P under the reflection in n. Since T is on the parabola, then $TP = TQ$ and T is on n. By the argument above, line n is a tangent to the parabola. Since at a point on a parabola there is only one tangent to the parabola, we must conclude that $n = t$ and that Q is the image of P under the reflection in t, as desired. ∎

Which folding operations do we want to incorporate into a definition of those points and lines that can be constructed by paperfolding? The answer to this question is by no means obvious. Unlike the other chapters, we cannot immediately write down a reasonable definition and go on from there. What follows below is the result of hindsight. If you have experimented with folding waxed-paper, then you know that many of the elementary constructions are possible. You should have thought to try verging. However, verging in one folding is generally impossible. Also, the rules of the game do not allow the folded paper to be folded again without first unfolding. We do insist that each postulated operation must be a single folding operation. Unlike most of our previous situations, we do not use a pencil here to record our constructions. The creases in the waxed-paper form a permanent record of our constructions.

Although at first it is not clear why, we will concentrate on only one *fundamental folding operation*: If for two given points P and Q and for given lines p and q there are only a finite number of lines t such that both P^t is on p and Q^t is on q, then we may fold the paper along each of these lines t. This models folding the waxed-paper along a line such that in the folding we have P falling on p and Q falling on q. We have omitted from all consideration the bothersome case $P = Q$. You should give yourself several configurations of P,p,Q,q and locate the lines t such that P^t is on p and Q^t is on q for each configuration. You do not need waxed-paper to try this. Draw points P and Q and lines p and q as in Figure 10.3 on a regular sheet of paper. Then go to a lighted window to fold the paper on the lines t that do the job. Several executions of this exercise will persuade you that we have hope in showing algebraically that the fundamental folding operation is reasonable. In pursuit of this goal, we next examine the fundamental folding operation in relation to all the possible cases of incidence among P, p, Q, and q, with $P \neq Q$. The next four paragraphs examine the different cases.

Suppose P on p, Q on q, and $p \parallel q$. Then we have P^t is on p and Q^t is on q for each perpendicular t to p. Since there are infinitely many lines t that give both P^t is on p and Q^t is on q, this first case is excluded by the finiteness condition of our fundamental folding operation. We do not want every point in the plane to be constructible by paperfolding. Although this case is something of a disaster, the remaining cases will be better behaved.

Suppose P on p, Q on q, and $p \nparallel q$. Here the solutions for lines t are the perpendicular from P to q, the perpendicular from Q to p, and the line joining P and Q. These lines number at most three and each is algebraically determined by an equation that is linear or quadratic.

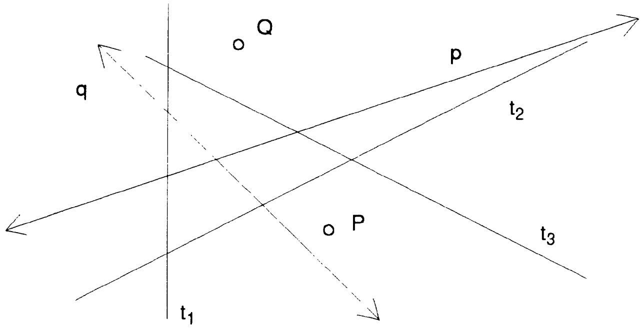


FIGURE 10.3.

Suppose P off p and $p \nparallel q$. Since all parabolas are similar (Exercise 10.6), we may suppose without loss of generality that $P = (0,0)$ and p has equation $Y = 2$. Line q has equation of the form $X + rY + s = 0$ since $q \nparallel p$. Let $Q = (u, v)$. Since line t cannot be perpendicular to p if P^t is to be on p, then t must have an equation of the form $Y = mX + b$. For use in Theorem 10.2, this equation for t is expressed in the form $mX - Y + b = 0$. So P^t on p then implies $y' = y - 2(-1)(mx - y + b)/(m^2 + (-1)^2)$ with $y' = 2$ and $x = y = 0$. This reduces to $b = m^2 + 1$. With $Q^t = (u', v')$, that Q^t is on q requires $u' + rv' + s = 0$. So, by Theorem 10.2, we have

$$\begin{aligned}&\{u - 2m[mu - v + (m^2 + 1)]/(m2 + 1)\}\\ &+ r\{v + 2[mu - v + (m^2 + 1)]/(m^2 + 1)\}\\ &+ s = 0.\end{aligned}$$

This reduces to a monic cubic in m with coefficients that are rational expressions in r,s,u,v. Fortunately(!), we don't have to solve the cubic, just know that there are always exactly 1, 2, or 3 solutions for m. A solution m uniquely determines b and so uniquely determines line t. In this case, line t is among the tangents to the parabola with focus P and directrix p. If Q is off q, then t is also among the tangents to the parabola with focus Q and directrix q. The case Q off q and $p \neq q$ is analogous to this case.

Suppose P off p and $p \parallel q$. Here, lines t are determined by quadratic or linear equations, and there are at most two solutions (Exercise 10.13). Analogous results hold for the case Q off q and $p \parallel q$.

We check back and confirm that all cases have been considered. Therefore, except when all three of $p \parallel q$, P on p, and Q on q hold simultaneously, there are at most three lines that give P^t on p and Q^t on q. Each of these lines is determined in the cartesian plane by a polynomial equation of degree at most 3 with coefficients from the smallest field containing the coordinates of P, Q, and the intercepts of p and q with coordinate axes. Since the marked ruler can be used to solve exactly those problems of polynomial degree 4 or less (Corollary 9.9), it follows that any point constructed by application of the fundamental folding operation alone can be constructed with a marked ruler, provided we are not given any unusual points to begin with. This result will be our Theorem 10.5, whose statement must wait until we have provided in Definition 10.4 the formal description of the points and lines that can be constructed by the fundamental folding operation.

In formulating the definition of those points and lines constructible by the fundamental folding operation, we see that only the one case where $p \parallel q$, P on p, and Q on q must be avoided. On the other hand, we might avoid the case $p \parallel q$ altogether. Not only will we do this, but it turns out that we are able to restrict the application of the fundamental folding operation to only the two special cases where $q = \overleftrightarrow{PQ}$ or $p \perp q$. Yates's postulate (iii), mentioned at the beginning of this chapter, is included in the case $q = \overleftrightarrow{PQ}$. The starter set must determine two points and two lines in order to apply the fundamental folding operation at least once. It takes at least three points to determine two lines. The three points $(0,0)$, $(1,0)$, $(0,1)$ determine the three lines having the equations $Y = 0$, $X = 0$, $X + Y = 1$; also, these three lines determine these three points.

Armed with some set of points as the starter set and following the format of our previous definitions to express the fundamental folding operation, we quickly run into language complications. Try it. The clever escape is to take the starter set as a set of lines and then define the constructible points in terms of the constructible lines. After all, the fundamental folding operation creates lines foremost. A newly created line on the waxed-paper, corresponding to a newly constructed line in the theory, is consistently denoted by t in this chapter, as a reminder that the lines are often the tangents described in Theorem 10.3. Now, provided you have carried out several applications of the fundamental folding operation on your own sheets of waxed-paper, the following definition easily models restricted applications of the fundamental folding operation. Theorem 10.5 follows from Definition 10.4 by the remarks above.

Definition 10.4. In the cartesian plane, a line is a ***paper line*** if the line is the last of a finite sequence $t_1, t_2, \ldots, t_n$ of lines such that each line has one of the three equations $Y = 0$, $X = 0$, $X + Y = 1$, or is a line t such that P^t is on p and Q^t is on q where: (i) p and q are nonparallel lines that appear earlier in the sequence; (ii) P and Q are distinct points each of which is the intersection of two lines that appear earlier in the sequence; and (iii) either $q = \overleftrightarrow{PQ}$ or $p \perp q$. A ***paper point*** is a point that is the intersection of two paper lines. A ***paper circle*** is a circle through a paper point with a paper point as center. A number x is a ***paper number*** if $(x, 0)$ is a paper point.

Theorem 10.5. *A paper point is a marked ruler point.*

Since none of the elementary constructions has been specifically modeled into the definition, there might be some apprehension that the definition is too narrow to model accurately all the physical constructions that are possible. That the converse of Theorem 10.5 is also true is perhaps surprising and is the consequence of a sequence of seven lemmas. The lemmas follow Theorem 10.6, which should be compared with Theorem 2.2.

Theorem 10.6. *If P and Q are distinct paper points, p and q are nonparallel paper lines such that $q = \overleftrightarrow{PQ}$ or $p \perp q$, and t is a line such that P^t is on p and Q^t is on q, then t is a paper line.*

Proof. Since p and q are paper lines, there are two sequences $p_1, p_2, \ldots, p$ and $q_1, q_2, \ldots, q$ of lines such that each sequence satisfies the condition of Definition 10.4. Since P is a paper point, there are two sequences $r_1, r_2, \ldots, r$ and $s_1, s_2, \ldots, s$ of lines such that each sequence satisfies the condition of Definition 10.4 and such that P is the intersection of r and s. Likewise, since Q is a paper point, there are two sequences $u_1, u_2, \ldots, u$ and $v_1, v_2, \ldots, v$ of lines such that each sequence satisfies the condition of Definition 10.4 and such that Q is the intersection of u and v. Now, the concatenate sequence

$$p_1, p_2, \ldots, p, q_1, q_2, \ldots, q, r_1, r_2, \ldots, r,$$
$$s_1, s_2, \ldots, s, u_1, u_2, \ldots, u, v_1, v_2, \ldots, v$$

satisfies the condition of Definition 10.4. Therefore, the sequence

$$p_1, p_2, \ldots, p, q_1, q_2, \ldots, q, r_1, r_2, \ldots, r,$$
$$s_1, s_2, \ldots, s, u_1, u_2, \ldots, u, v_1, v_2, \ldots, v, t$$

must also satisfy the condition of Definition 10.4. Thus, t is a paper line by Definition 10.4. ∎

Lemma 10.7. *Every paper point is on two paper lines. Points* $(1, 0)$, $(0, 1)$, $(2, 0)$, *and* $(0, 2)$ *are paper points. Every paper line passes through two paper points.*

Proof. The points $(0, 0)$, $(1, 0)$, and $(0, 1)$ are paper points since each lies on two of the lines having one of the three equations $Y = 0$, $X = 0$, $X+Y = 1$. Since any other paper point must lie on two paper lines by Definition 10.4, then every paper point is on two paper lines. If $P = (1, 0)$, $Q = (0, 0)$, p has equation $X + Y = 1$, and $q = \overleftrightarrow{PQ}$, then P^t is on p and Q^t is on q when t has the equation $Y = X$ or $X = 1$. Hence, the two lines with these equations are paper lines by Theorem 10.6. So $(1, 1)$ is a paper point. Next, if $P = (1, 1)$, $Q = (0, 0)$, p has equation $X = 1$, and $q = \overleftrightarrow{PQ}$, then P^t is on p and Q^t is on q when t has the equation $X + Y = 2$. Hence the line with equation $X + Y = 2$ is a paper line by Theorem 10.6. So $(2, 0)$ and $(0, 2)$ are paper points. Finally, since every paper line intersects the parallel paper lines with equations $X = 0$ and $X = 1$ in two paper points or else intersects the parallel paper lines with equations $X+Y = 1$ and $X+Y = 2$ in two paper points, then every paper line contains two paper points. ∎

The paper points mentioned in the preceding lemma make up the starter set for the ruler points. We need the converse of the last part of the lemma to show that we have the power of the ruler. The desired converse is Lemma 10.9, whose proof requires the next lemma.

Lemma 10.8 (*cf.* Euclid I.11 and I.12). *The line perpendicular to a given paper line and through a given paper point is a paper line.*

Proof. Suppose P is a paper point on paper line q and that line t is the perpendicular to q at P. By Lemma 10.7, there is a paper line p on P that is different from q and there is a paper point Q on q that is different from P. Then $q = \overleftrightarrow{PQ}$, P^t is on p, and Q^t is on q. Therefore, t is a paper line in this case by Theorem 10.6.

Now, suppose Q is a paper point off paper line p and that t is the perpendicular to p that passes through Q. By Lemma 10.7, there is a paper line q through Q that intersects p, since there are two paper lines through Q and at most one can be parallel to p. So P, the intersection of p and q, is a paper point, and $q = \overleftrightarrow{PQ}$. Since P^t is on p and Q^t is on q, then t is a paper line in this case by Theorem 10.6. ■

Constructing the perpendicular through a given point to a given line is undoubtedly one of the elementary constructions you devised on your own. To construct the line t, without constructing p and Q as they appear in the proofs above, you need only fold the waxed-paper so that q falls on itself and the crease passes through P. You should feel free to use this shortcut now that Lemma 10.8 has been proved.

Lemma 10.9. *Any two paper points are on a paper line.*

Proof. Suppose P and Q are two paper points. There is a paper line p through P by Lemma 10.7. There is a paper line q through Q that is perpendicular to p by Lemma 10.8. Since $p \perp q$, P^t is on p, and Q^t is on q where $t = \overleftrightarrow{PQ}$, then t is a paper line by Theorem 10.6. ■

The lines p and q in the proof above are useless for folding the line through two points on waxed-paper. The lines are necessary for the theory to show that such a folding is permissible. In illustrating a paperfolding construction on waxed-paper, we would simply fold on $\overleftrightarrow{PQ}$ and not introduce the lines p and q from the proof into the illustration. Yates's postulate (ii) is not explicitly written into our Definition 10.4 because it is implicit in the fundamental folding operation by Lemma 10.9. With Lemmas 10.7 and 10.9, we now have the power of a ruler. Next, we head for a cannon.

Lemma 10.10. *If P and Q are two paper points and paper line p passes through Q, then the intersections of p and Q_P are paper points.*

Proof. With q the paper line through paper points P and Q and with p a paper line through Q, Theorem 10.6 gives the angle bisectors of the angles determined by p and q as paper lines. Since the perpendiculars from P to the angle bisectors intersect p at paper points the same distance from Q as P, the lemma follows. ■

Within the proof of Lemma 10.10 there are paperfolding constructions for angle bisectors. On a sheet of waxed-paper, the angle bisector of $\angle AVB$ is obtained by folding $\overrightarrow{VA}$ and $\overrightarrow{VB}$ together. Also, with P,p,Q,q as in the proof above, the perpendicular bisector of $\overline{PQ}$ is obtained in the one folding that brings P and Q together. This is, of course, right out of the definition of a reflection. This fundamental folding operation is Yates's postulate (i). The midpoint of P and Q is then easily constructed, as well. As a result of Lemma 10.10, we have the power of ruler and dividers. We next head for a poncelet circle with some center Q.

Lemma 10.11. *If P and Q are two paper points and p is a paper line, then any point of intersection of p and Q_P is a paper point.*

Proof. If p is on Q, we are done by the preceding lemma. Suppose p is off Q and intersects Q_P at a point R. We want to show that R is a paper point. Let $q = \overleftrightarrow{PQ}$. We may suppose $p \nparallel q$ without loss of generality. Then, by Theorem 10.6, the perpendicular bisector t of $\overline{PR}$ is a paper line. Since the perpendicular from P to t is a paper line by Lemma 10.8 and intersects p at R, then R is itself a paper point. ■

It now follows that all ruler and compass points are paper points. With Lemma 10.9 we were up to constructions with ruler alone and thus to the points with coordinates in $\mathbb{Q}$. Lemma 10.10 took us to constructions by ruler and dividers and thus to the points with coordinates in $\mathbb{P}$. By Lemma 10.11 and the Steiner–Poncelet Theorem, we now have all the points with coordinates in $\mathbb{E}$. Next, we consider angle trisection and then cube roots. By the results of Chapter 9, success in this will bring us to all the points with coordinates in $\mathbb{V}$.

Lemma 10.12. *The angle trisectors of the angles formed by intersecting paper lines are themselves paper lines.*

Proof. Suppose P,Q,R are paper points with $\angle PQR$ acute. Let M be the midpoint of $\overline{PQ}$, let p be the perpendicular from M to $\overleftrightarrow{QR}$, and let q be the perpendicular to p at M. Of the three lines t with P^t on p and Q^t on q, let t be the one intersecting $\overline{PM}$. As in Figure 10.4, let $S = Q^t$ and $T = P^t$. Let $\overrightarrow{QS}$ intersect p at U. Then $\triangle PMT \cong \triangle QMU$ by ASA. So $TM = MU$ and $\triangle TMS \cong \triangle UMS$ by SAS. Hence $\angle TSM$ is congruent to $\angle MSQ$, which is congruent to $\angle SQR$. Since $\angle TSQ \cong \angle PQS$ by the Pons Asinorum, then $\overrightarrow{QS}$ trisects $\angle PQR$, as desired. (If t and q intersect at V, then $\overrightarrow{QV}$ is the other trisector of $\angle PQR$.) The result follows for obtuse angles angles as well because trisecting a right angle is a ruler and compass construction. (However, see Exercise 10.11.) ■

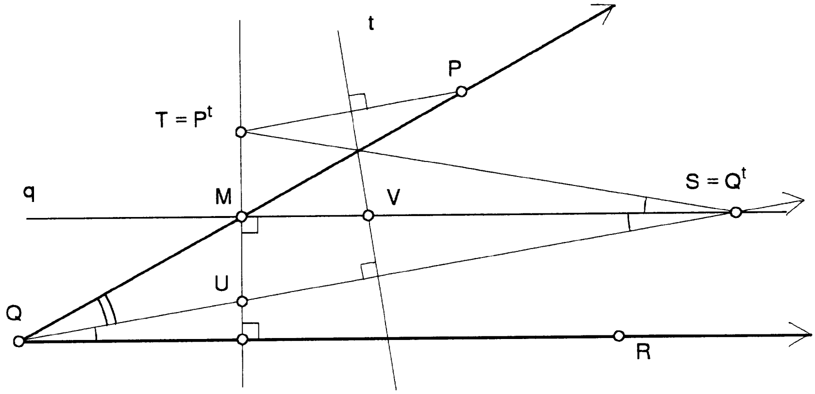


FIGURE 10.4.

Since $2MQ = PQ = TS = SU$ in Figure 10.4, this angle trisection is related to the trisection construction of Pappus that uses verging and was given in Theorem 9.2. Dayoub and Lott gave the construction in the proof of the lemma in 1977 for the Mira, which is a transparent plastic reflector invented by Scroggie and Gillespie and distributed in the United States by Dale Seymour Publications (800-872-1100; P.O. Box 10888, Palo Alto, CA 94303) and by Creative Publications, (800-624-0822; 5623 W 115th Street, Worth, IL 60482). A Mira is shown in Figure 10.5. Invented for use in the study of transformation geometry, the Mira is an excellent construction tool, although somewhat expensive. The construction theory for the Mira is the same as that for paperfolding. However, you should get your hands on one for the delight of making some constructions with this marvelous tool.

A related but different construction for trisecting an angle by paperfolding can be found in Exercise 10.14. This elegant construction is due to Hisashi Abe and was published in Japanese in 1980.

The field of paper numbers is closed under trisection. We know from Vieta's result that is "very worth knowing" that we are only one step away from showing that the marked ruler points are paper points. We need to show only that the paper numbers are closed under cube root. All attempts

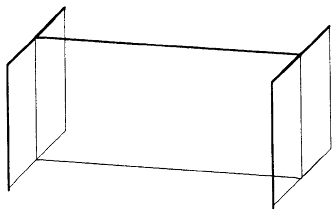


FIGURE 10.5.

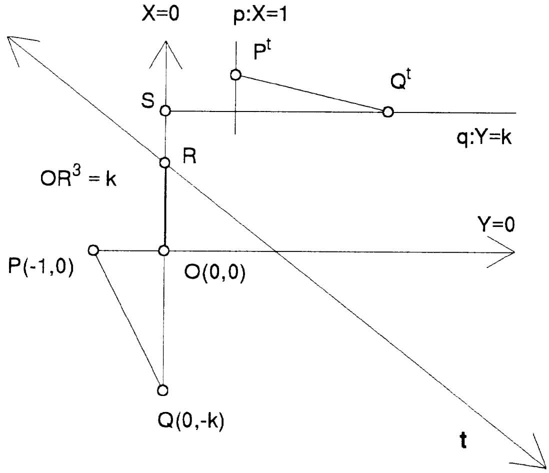


FIGURE 10.6.

at bending the classical verging solutions by Nicomedes, Apollonius, Vieta, and Newton to the problem of the cube root by paperfolding have apparently failed. Nevertheless, the following construction, published by this author in 1979, does do the job. Figure 10.6 emphasizes the use of the fundamental folding operation in the construction.

Lemma 10.13. *If O and S are paper points such that $OS = k$, then there is a paper point R on $\overline{OS}$ such that $OR = \sqrt[3]{k}$.*

Proof. Without loss of generality, we may suppose $O = (0,0)$ and $S = (0,k)$. Let $P = (-1,0)$ and $Q = (0,-k)$; P and Q are paper points. Let p have equation $X = 1$ and let q have equation $Y = k$; p and q are paper lines. See Figure 10.7 and recall Theorem 10.3. Since two parabolas with

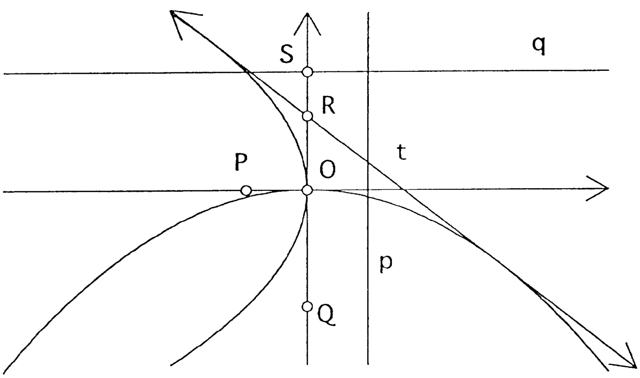


FIGURE 10.7.

a common vertex and perpendicular axes have a unique common tangent, there is a unique line t such that P^t is on p and Q^t is on q. By the definition, t is a paper line and intersects the Y-axis at a paper point R. Exercise 10.9 requires you to show that $R = (0, \sqrt[3]{k})$. Thus, R is a paper point such that $OR = \sqrt[3]{k}$. ∎

Since any marked ruler construction can be reduced to a sequence of quadratic problems, trisection problems, and cube root problems, our sequence of lemmas has shown that any marked ruler point is a paper point. We have the converse of Theorem 10.5 and the principal theorem of paperfolding constructions.

Theorem 10.14. *A point is a paper point iff the point is a marked ruler point. The field of paper numbers is* $\mathbb{V}$.

For given point P and given line m, we know from our theory that P^m can be constructed by applying the fundamental folding operation alone. Exercise 10.1 asks that you do exactly that. In practice, however, the paperfolding constructions are often done with a pencil in hand to mark the image of a given point P under the reflection in line m when the paper is folded along m. This tremendous short cut in constructing P^m eliminates doing elementary paperfolding constructions over and over again. This is analogous to using a modern compass as a dividers when illustrating ruler and compass constructions, once Euclid I.3 has been proved. There is also an analogue to using a plastic right-angle in doing euclidean construction to eliminate the tedium of constructing perpendicular lines. The name of the game is to make the constructions interesting but not to be bored by too much repetition. You might want to consider the consequences of allowing more that one fold at a time. To fold the paper when the paper is already folded, which we have not allowed previously, presents possibilities for exploration. Again, you have to decide what game you want to play. The important thing is to enjoy the game.

Exercises

10.1. Given point P and line m, give a paperfolding construction for P^m.◊

10.2. Give a paperfolding construction for a square having given side $\overline{AB}$.◊

10.3. Give a paperfolding construction corresponding to Euclid I.3.◊

10.4. Give a paperfolding construction corresponding to Euclid I.1.◊

10.5. In three separate illustrations, each like Figure 10.3 and each with a different incidence of P,Q,p,q on waxed-paper, illustrate the construction by paperfolding of all the lines t determined by the fundamental folding operation.

10.6. Show that all parabolas are similar.◊

10.7. State at least three theorems of elementary geometry suggested by folding a triangle, cut out from any kind of paper, along the three dotted lines shown in Figure 10.8.◊

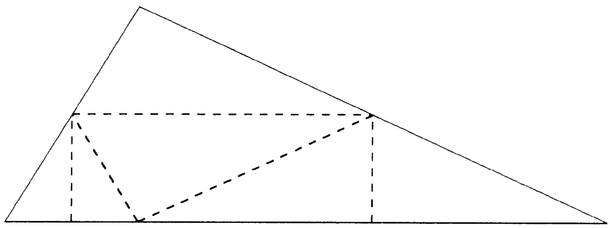


FIGURE 10.8.

10.8. Place point F in the center of a sheet of lined paper. For each line on the paper, as in Figure 10.9, fold so that the end of the line is at F and mark the intersection of the line with that crease. What do you have?

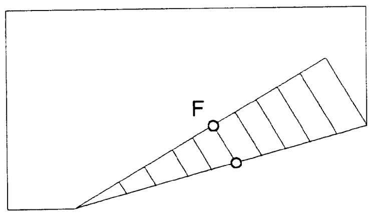


FIGURE 10.9.

10.9. Do the algebra necessary to finish the proof of Lemma 10.13.◊

10.10. Obtain the formulas given in the statement of Theorem 10.2 from the equations in the proof.◊

10.11. Give a paperfolding construction for trisecting a right angle. Construct by paperfolding an illustration of the angle trisection construction of Lemma 10.12.◊

10.12. Prove: If P and Q are distinct paper points, p and q are nonparallel paper lines, and t is a line such that P^t is on p and Q^t is on q, then t is a paper line.

10.13. Show that if P is off p and $p \parallel q$, then for any point Q different from P there are at most two lines t such that P^t is on p and Q^t is on q.◊

10.14. Suppose $\angle PQR$ is acute with M the midpoint of $\overline{PQ}$. Let l be the perpendicular to $\overline{QR}$ at Q. Let N be the foot of the perpendicular from P

to l. Let $n = \overleftrightarrow{PQ}$. Let q be the line through M that is perpendicular to l. Let t be the line intersecting $\overline{PM}$ such that N^t is on n and such that Q^t is on q. Let $S = Q^t$. Let t and q intersect at V. Show that $\overrightarrow{QS}$ and $\overrightarrow{QV}$ trisect $\angle PQR$.◊

10.15. Suppose paper square $\square ABCD$ is trisected by $\overline{PQ}$ and by $\overline{RS}$ as in Figure 10.10. With line t such that C^t is on $\overline{AB}$ and s^t is on $\overline{PQ}$, let $T = C^t$ and find AT/TB.◊

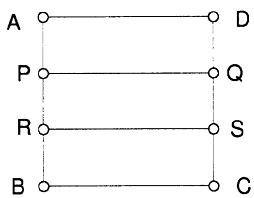


FIGURE 10.10.

10.16. Starting with a given circle C_B drawn on a sheet of waxed-paper and with a given point F inside/outside the circle, what is the locus of all points X such that t and $\overline{CP}$ intersect at X where $P = F^t$ and $CP = CB$?◊

10.17. Given paper points F and P and given paper line d, then show that a line through P and normal to the parabola with focus F and directrix d is a paper line.◊

10.18. Use paperfolding constructions to illustrate the following theorems: The medians of a triangle are concurrent at a point G, called the *centroid* of the triangle. The perpendicular bisectors of the sides of a triangle are concurrent at a point O, called the *circumcenter* of the triangle. The altitudes of a triangle are concurrent at a point H, the *orthocenter* of the triangle. Further, points H, G, and O are collinear, on a line called the *Euler line* when the points are distinct. In fact, G trisects $\overline{HO}$ unless $G = H = O$.

Suggested Reading and References

For further reading, consider *College Geometry* by Nathan Altshiller Court and *Ruler and Compasses* by Hilda P. Hudson. These books have been reprinted in inexpensive editions by Chelsea Publishing Company. If you read German, seek out the book by L. Bieberbach; if you read French, seek out the book by Henri Lebesgue. Also, you may find a lot of interesting geometry in *Foundations of Geometry* by David Hilbert. For biographies of mathematicians and scientists, enjoy wandering among the eighteen volumes of the *Dictionary of Scientific Biography*, edited by Charles Coulston Gillispie.

The following list of references includes only those references on geometric constructions that the author has seen. For each entry, the name of the author is as it appears on the cited work. A complete bibliography would be much longer than this entire book. Many more references to the subject, especially references at an elementary level, may be found in various bibliographies by William L. Schaaf.

Aaboe, Asger, *Episodes from the Early History of Mathematics*, New Mathematical Library Vol. 13, Mathematical Association of America, 1964.

Adler, August, *Theorie der Geometrischen Konstruktionen*, Göschen, 1906.

Altshiller Court, Nathan, *College Geometry*, Barnes and Noble, 1952.

Altshiller Court, Nathan, *See* Court, N. A.

Archibald, R. C., Constructions with a double-edged ruler, *American Mathematical Monthly* 25 (1918) 358–360.

Archibald, R. C., Gauss and the regular polygon of seventeen sides, *American Mathematical Monthly* 27 (1920) 323–326.

Archibald, R. C., Malfatti's problem, *Scripta Mathematica* 1 (1932) 170–171.

Ball, W. W. Rouse and H. S. M. Coxeter, *Mathematical Recreations & Essays* Twelth Edition, Toronto University Press, 1974.

Becker, Jerry P., On solutions of geometrical constructions utilizing the compasses alone, *Mathematics Teacher* 57 (1964) 398–403.

Bell, E. T., *The Development of Mathematics*, McGraw-Hill, 1940.

Berggren, J. L., *Episodes in the Mathematics of Medieval Islam*, Springer-Verlag, 1986.

Bieberbach, L., *Theorie der geometrischen Konstruktionen*, Birkhaüser, 1952.

Birkhoff, Garrett and Saunders MacLane, *A Survey of Modern Algebra* Third Edition, Macmillan, 1965.

Bishop, Wayne, How to construct a regular polygon, *American Mathematical Monthly* 85 (1978) 186–188.

Bold, Benjamin, *Famous Problems of Geometry and How to Solve Them*, Dover (reprint), 1982.

Bos, H. J. M., Arguments on motivation in the rise and decline of a mathematical theory; the "construction of equations", 1632–ca.1750, *Archive for History of Exact Sciences* 30 (1984) 331–380.

Boyer, Carl B., *A History of Mathematics*, Wiley, 1968.

Breidenbach, Walter, *Die Dreiteilung des Winkels*, Teubner, 1951.

Breidenbach, Walter, *Das Delische Problem*, Teubner, 1953.

Breidenbach, W. and W. Süss, Geometric constructions IN *Fundamentals of Mathematics* Vol. 2 (*Geometry*), (H. Behnke, F. Bachmann, K. Fladt, and H. Kunle, eds.), pp 198–237, MIT, 1974.

Brocard, H., Note sur la division mécanique de l'angle, *Bulletin de la Société Mathématique de France* 5 (1877) 43–47.

Bunt, Lucas N. H., Philip S. Jones, and Jack D. Bedient, *The Historical Roots of Elementary Mathematics*, Prentice Hall, 1976.

Bussey, W. H., Geometric constructions without the classical restriction to ruler and compasses, *American Mathematical Monthly* 43 (1936) 265–280.

Cajori, Florian, A forerunner of Mascheroni, *American Mathematical Monthly* 36 (1929) 364–365.

Cannon, Larry, Geometry of the rational plane, *College Mathematics Journal* 17 (1986) 392–402.

Carnahan, Walter H., Compass geometry, *School Science and Mathematics* 32 (1932) 384–390.

Carnahan, Walter H., Geometrical constructions without the compass, *School Science and Mathematics* 36 (1936) 182–189.

Cheney, Wm. Fitch, Jr., Can we outdo Mascheroni?, *Mathematics Teacher* 46 (1953) 152–156.

Coolidge, Julian Lowell, *A History of Geometric Methods*, Dover (reprint), 1963.

Coolidge, Julian Lowell, *A Treatise on the Circle and the Sphere*, Chelsea (reprint), 1971.

Courant, Richard and Herbert Robbins, *What Is Mathematics?*, Oxford, 1941.

Court, N. A., Mascheroni constructions, *Mathematics Teacher* 51 (1958) 370–372.

Court, N. A., The problem of Apollonius, *Mathematics Teacher* 54 (1961) 444–452.

Court, N. A., *See* Altshiller Court, Nathan.

Croft, H. T. and T. W. Körner, Constructions with a rigid compass, *Mathematical Gazette* 62 (1978) 85–89.

Dantzig, Tobias, *The Bequest of the Greeks*, Scribner, 1955.

Davis, David R., *Modern College Geometry*, Addison-Wesley, 1949

Dawson, T. R., "Match-stick" geometry, *Mathematical Gazette* 23 (1939) 161–168.

Dayoub, I. M. and J. W. Lott, What can be done with a Mira?, *Mathematics Teacher* 70 (1977) 394–399.

Dedron, P. and J. Itard, *Mathematics and Mathematicians* (2 vols.), Transworld, 1974.

Dennis, David, René Descartes' curve-drawing devices: experiments in relations between mechanical motion and symbolic language, *Mathematics Magazine* 70 (1997) 163–174.

DeTemple, Duane W., Simple constructions for the regular pentagon and heptadecagon, *Mathematics Teacher* 82 (1982) 361–365.

Dickson, Leonard Eugene, *New First Course in the Theory of Equations*, Wiley, 1939.

Dickson, L. E., Constructions with Ruler and Compasses; Regular Polygons IN *Monographs on Topics of Modern Mathematics Relevant to the Elementary Field* (J. W. A. Young, ed.), Dover (reprint), 1955.

Dörrie, Heinrich, *100 Great Problems of Elementary Mathematics*, Dover, 1965.

Dudley, Underwood, What to do when the trisector comes, *Mathematical Intelligencer*, 5 (1983) 20–25.

Dudley, Underwood, *A Budget of Trisectors*, Springer-Verlag, 1987.

Duncan, Donald and William Barnier, On trisection, quintisection, ... , etc., *American Mathematical Monthly* 89 (1982) 693.

Eidswick, Jack, Two trisectrices for the price of one rolling coin, *College Mathematics Journal* 24 (1993) 422–430.

Emert, John W., Kay I. Meeks, and Roger B. Nelson, Reflections on a Mira, *American Mathematical Monthly* 101 (1994) 544–549.

Eves, Howard, *An Introduction to the History of Mathematics*, Rinehart, 1953.

Eves, Howard, *A Survey of Geometry* (2 vols.), Allyn and Bacon, 1965.

Euclid, *Elements*, *See* Heath, Sir Thomas L.

Fraleigh, John B., *A First Course in Abstract Algebra*, Addison-Wesley, 1967.

Gardner, Randolph Scott, *Instruments for the Enrichment of Secondary School Mathematics*, Edwards, 1951.

Geiselmann, Harrison A., Have you tried this? An alternate construction of a regular pentagon, *New York State Mathematics Teachers' Journal* 31 (1981) 56–58.

Gertschläger, Robert, Euclidean constructions and the geometry of origami, *Mathematics Magazine* 68 (1995) 357–371.

Gibb, Allan A., There are more ways than one to bisect a segment, *Mathematics Teacher* 70 (1977) 390–393.

Gleason, Andrew M., Angle trisection, the heptagon, and the triskaidecagon, *American Mathematical Monthly* 95 (1988) 185–194.

Glotin, De quelques moyens pratiques de diviser les angles en parties égales, *Mémoires de la Société des Sciences Physiques et Naturelles de Bordeaux* 2 (1863) 253–278 and plates.

Goldberg, Michael, On the original Malfatti problem, *Mathematics Magazine* 40 (1967) 241–247.

Goldberg, Michael, The converse Malfatti problem, *Mathematics Magazine* 41 (1968) 262–266.

Hadlock, Charles Robert, *Field Theory and its Classical Problems*, Carus Mathematical Monographs Vol. 19, Mathematical Association of America, 1978.

Hallerberg, Arthur E., The geometry of the fixed compass, *Mathematics Teacher* 52 (1959) 230–244.

Hallerberg, Arthur E., Georg Mohr and Euclidis Curiosi, *Mathematics Teacher* 53 (1960) 127–132

Heath, Sir Thomas L., *A History of Greek Mathematics*, Oxford, 1921.

Heath, Sir Thomas L., *The Thirteen Books of Euclid's Elements* (3 vols.), Dover (reprint), 1956.

Hess, Adrien L., Certain topics related to constructions with straightedge and compasses, *Mathematics Magazine* 29 (1956) 217–221.

Hilbert, David, *Foundations of Geometry* Tenth Edition (revised and enlarged by Paul Bernays), Open Court, 1971.

Hlavaty, Julius H., Mascheroni constructions, *Mathematics Teacher* 50 (1957) 482–487.

Hobson, E. W., "Squaring the Circle" A History of the Problem, IN *Squaring the Circle and Other Monographs*, Chelsea (reprint), 1969.

Hogendijk, Jan P., Greek and Arabic constructions of the regular heptagon, *Archive for History of Exact Sciences* 30 (1984) 197–330.

Hogendijk, Jan P., Arabic traces of lost works of Apollonius, *Archive for History of Exact Sciences* 35 (1986) 187–253.

Honsberger, Ross, *Mathematical Gems*, Dolciani Mathematical Expositions Vol. 1, Mathematical Association of America, 1973.

Honsberger, Ross, *Ingenuity in Mathematics*, New Mathematical Library Vol. 23, Mathematical Association of America, 1977.

Honsberger, Ross, *Mathematical Morsels*, Dolciani Mathematical Expositions Vol. 3, Mathematical Association of America, 1978.

Hudson, Hilda P., Ruler and Compasses IN *Squaring the Circle and Other Monographs*, Chelsea (reprint), 1969.

Hull, Thomas, A note on "impossible" paper folding, *American Mathematical Monthly* 103 (1996) 240–241.

Huzita, H., The trisection of a given angle solved by the geometry of origami IN *Proceedings of the First International Meeting of Oragami Science and Technology* (Ferra, Italy, December 1989; Humiaki Huzita, ed.).

Huzita, Humiaki, ed., *Proceedings of the First International Meeting of Oragami Science and Technology* (Ferra, Italy, December 1989), no date.

Huzita, Humiaki, Drawing the regular heptagon and the regular nonagon by origami (paper folding), *Symmetry: Culture and Science* 5 (1994) 69–83.

Ibn al-Haytham, *On the Quadrature of the Circle. See* Suter, Heinrich.

Isaacs, I. M., Solution of polynomials by real radicals, *American Mathematical Monthly* 92 (1985) 571–575.

Jacobson, Nathan, *Basic Algebra I*, Freeman, 1974.

Johnson, Donovan A., *Paper Folding for the Mathematics Class*, National Council of Teachers of Mathematics, 1957.

Johnson, Roger A., *Advanced Euclidean Geometry (Modern Geometry)*, Dover (reprint), 1960.

Juster, Norton, *The Dot and the Line*, Random House, 1963.

Kazarinoff, Nicholas D., *Ruler and the Round*, Prindle, Weber, and Schmidt, 1970.

Kijne, D., Construction geometries and construction fields IN *Algebraic and Topological Foundations of Geometry* (H. Freudenthal, ed.), pp 87–97, Pergamon, 1962.

Klein, Felix, *Famous Problems of Elementary Geometry*, Dover (reprint), 1956.

Kline, Morris, *Mathematical Thought from Ancient to Modern Times*, Oxford, 1972.

Kostovskii, A. N., *Geometrical Constructions Using Compasses Only*, Blaisdell, 1961.

Kramer, Edna E., *The Nature and Growth of Modern Mathematics*, Princeton, 1970.

Kulczycki, Stefan, *Non-Euclidean Geometry*, Pergamon, 1961.

Kutuzov, B. V., *Geometry*, Studies in Mathematics Vol. IV, School Mathematics Study Group, 1960.

Lebesgue, Henri, *Leçons sur les Constructions Géométriques*, Gauthier-Villars, Paris 1950.

Lob, H. and H. W. Richmond, On the solution of Malfatti's problem for a triangle, *Proceedings of the London Mathematical Society* Series 2, 30 (1930) 287–304.

Mahoney, Michael Sean, *The Mathematical Career of Pierre de Fermat*, Princeton, 1973.

Martin, George E., Duplicating the cube with a mira, *Mathematics Teacher* 72 (1979) 204–208.

Martin, George E., Paper-folding, *New York State Mathematics Teachers' Journal* 35 (1985) 137–140.

Martin, George E., Toothpick constructions, Part 1, *New York State Mathematics Teachers' Journal* 36 (1986) 105–111.

Martin, George E., Toothpick constructions, Part 2, *New York State Mathematics Teachers' Journal* 36 (1986) 173–175.

Meschkowski, Herbert, *Unsolved and Unsolvable Problems in Geometry*, Oliver and Boyd, 1966.

Messer, Peter, Problem 1054; solution by Stanley Rabinowitz, *Crux Mathematicorum* 12 (1986) 284–285.

Mills, C. N., The problem of Napoleon, *Mathematics Teacher* 46 (1953) 344-345.

Moritz, Robert Edouard, *On Mathematics, A Collection of Witty, Profound, Amusing Passages about Mathematics and Mathematicians*, Dover (reprint), 1958.

Neumann, Peter M., Reflection in a spherical mirror, *American Mathematical Monthly*, in press.

Nivin, Ivan, *Irrational Numbers*, Carus Mathematical Monographs Vol. 11, Mathematical Association of America, 1956.

Olson, Alton T., *Mathematics Through Paper Folding*, National Council of Teachers of Mathematics, 1975.

Pedoe, D., *A Course of Geometry for Colleges and Universities*, Cambridge, 1970.

Pedoe, Dan, *Geometry and the Liberal Arts*, St. Martin's, 1976.

Pedoe, D., *Circles, A Mathematical View*, Dover (reprint), 1979.

Petersen, Julius, Methods and Theories for the Solution of Problems of Geometric Constructions IN *String Figures and Other Monographs*, Chelsea (reprint), 1960.

Pierpont, James, On an undemonstrated theorem of the Disquisitiones Arithmeticæ, *American Mathematical Society Bulletin* 2 (1895–1896) 77–83.

Plemelj, Josef, Die Siebenteilung des Kreises, *Monatshefte Mathematik und Physik* 23 (1912) 309–311.

Rademacher, Hans and Otto Toeplitz, *The Enjoyment of Mathematics*, Princeton, 1957.

Richmond, Herbert W., A construction for a regular polygon of seventeen sides, *Quarterly Journal of Pure and Applied Mathematics* 26 (1893) 206–207 and plate II.

Richmond, H. W., To construct a regular polygon of 17 sides, *Mathematische Annalen* 67 (1909) 459–461.

Ribenboim, P., Pythagorean and Fermatian fields, *Symposia Mathematica* 15 (1975) 583–593.

Robertson, Jack M., Geometric constructions using hinged mirrors, *Mathematics Teacher* 79 (1986) 380–386.

Rosen, Michael I., Niels Hendrik Able and equations of fifth degree, *American Mathematical Monthly* 102 (1995) 495–505.

Row, Sundara T., *See* Sundara Row, T.

Sabra, A. I., Ibn al-Haytham's lemmas for solving "Alhazen's problem", *Archive for History of Exact Sciences* 26 (1984) 299–324.

Seidenberg, A., Remarks on Nicomedes' duplication, *Archive for History of Exact Sciences* 3 (1966) 97–101.

Smith, David Eugene, The History and Transcendence of π IN *Monographs on Topics of Modern Mathematics Relevant to the Elementary Field* (J. W. A. Young, ed.), Dover (reprint), 1955.

Smith, John D., The remarkable Iban al-Haytham, *Mathematical Gazette* 76 (1992) 189–198.

Smogorzhevskii, A. S., *The Ruler in Geometrical Constructions*, Blaisdell, 1961.

Starke, E. P., The regular triacontagon, *American Mathematical Monthly* 51 (1944) 47–48.

Steiner, Jacob, *Geometrical Constructions with a Ruler Given a Fixed Circle with Its Center*, Scripta Mathematica, 1950.

Steinhaus, H., *Mathematical Snapshots* Third American Edition, Oxford, 1969.

Stewart, Ian, Gauss, *Scientific American* 237 (July 1977) 122–131.

Sundara Row, T., *Geometric Exercises in Paperfolding*, Dover (reprint), 1966.

Suter, Heinrich, Die Kreisquadratur des Ibn el Haiṯam, *Zeitschrift für Mathematik und Physik* Historisch-litterarische Abteilung 44 (1899) 33–47.

Tietz, Heinrich, *Famous Problems of Mathematics*, Graylock, 1965.

Trigg, Charles W., Unothhodox ways to trisect a line segment, *School Science and Mathematics* 54 (1954) 525–528.

van der Waerden, B. L., *Modern Algebra* Vol. 1, Ungar, 1953.

Videla, Carlos R., On points constructible from conics. *Mathematical Intelligencer* 19 (1997) 53–57.

Viéte, François, *The Analytic Art* (translated by T. Richard Witmer), Kent State, 1983.

Wantzel, L., Recherches sur les moyens de reconnaître si un probléme de géométrie peut se résoudre avec la régle et le compas, *Journal de Mathématiques Pures et Appliquées* 2 (1837) 366–372.

Wernick, William, Geometric construction: the double straightedge, *Mathematics Teacher* 64 (1971) 697–704.

Wills, Herbert, *Leonardo's Dessert, No π*, National Council of Teachers of Mathematics, 1985.

Yates, Robert C., Paper folding IN *Multi-Sensory Aids in the Teaching of Mathematics*, Eighteenth Yearbook, pp 154–159, National Council of Teachers of Mathematics, 1945.

Yates, Robert C., *Geometrical Tools, A Mathematical Sketch and Model Book*, Educational Publishers, 1949.

Yates, Robert C., Euclidean constructions, *Mathematics Teacher* 47 (1954) 231–233.

Yates, Robert C., *The Trisection Problem*, Classics in Mathematics Education Vol. 3, National Council of Teachers of Mathematics, 1971.

Zalgaller, V. A. and G. A. Los′, The solution to Malfatti's problem, *Journal of Mathematical Sciences* 72 (1994) 3163–3177.

Index

"It's *very* provoking," Humpty Dumpty said after a long silence, looking away from Alice as he spoke, "to be called an egg,—*very!*" . . .
"You've been listening at doors—and behind trees—and down chimneys—or you couldn't have know it!"
"I haven't, indeed!" Alice said very gently. "It's in a book."
"Ah, well! They may write such things in a *book*," Humpty Dumpty said in a calmer tone. . . .
"If I'd meant that, I'd have said it," said Humpty Dumpty . . .
"When *I* use a word," Humpty Dumpty said in a rather scornful tone, "it means just what I choose it to mean—neither more nor less."

Through the Looking Glass
LEWIS CARROLL
(Oxford mathematics teacher
Rev. Charles Lutwidge Dodgson)

Undergraduate Texts in Mathematics

(continued from page ii)

James: Topological and Uniform Spaces.
Jänich: Linear Algebra.
Jänich: Topology.
Kemeny/Snell: Finite Markov Chains.
Kinsey: Topology of Surfaces.
Klambauer: Aspects of Calculus.
Lang: A First Course in Calculus. Fifth edition.
Lang: Calculus of Several Variables. Third edition.
Lang: Introduction to Linear Algebra. Second edition.
Lang: Linear Algebra. Third edition.
Lang: Undergraduate Algebra. Second edition.
Lang: Undergraduate Analysis.
Lax/Burstein/Lax: Calculus with Applications and Computing. Volume 1.
LeCuyer: College Mathematics with APL.
Lidl/Pilz: Applied Abstract Algebra.
Macki-Strauss: Introduction to Optimal Control Theory.
Malitz: Introduction to Mathematical Logic.
Marsden/Weinstein: Calculus I, II, III. Second edition.
Martin: The Foundations of Geometry and the Non-Euclidean Plane.
Martin: Geometric Constructions.
Martin: Transformation Geometry: An Introduction to Symmetry.
Millman/Parker: Geometry: A Metric Approach with Models. Second edition.
Moschovakis: Notes on Set Theory.
Owen: A First Course in the Mathematical Foundations of Thermodynamics.
Palka: An Introduction to Complex Function Theory.
Pedrick: A First Course in Analysis.
Peressini/Sullivan/Uhl: The Mathematics of Nonlinear Programming.
Prenowitz/Jantosciak: Join Geometries.
Priestley: Calculus: An Historical Approach.
Protter/Morrey: A First Course in Real Analysis. Second edition.
Protter/Morrey: Intermediate Calculus. Second edition.
Roman: An Introduction to Coding and Information Theory.
Ross: Elementary Analysis: The Theory of Calculus.
Samuel: Projective Geometry. *Readings in Mathematics.*
Scharlau/Opolka: From Fermat to Minkowski.
Sethuraman: Rings, Fields, and Vector Spaces: An Approach to Geometric Constructability.
Sigler: Algebra.
Silverman/Tate: Rational Points on Elliptic Curves.
Simmonds: A Brief on Tensor Analysis. Second edition.
Singer: Geometry: Plane and Fancy.
Singer/Thorpe: Lecture Notes on Elementary Topology and Geometry.
Smith: Linear Algebra. Second edition.
Smith: Primer of Modern Analysis. Second edition.
Stanton/White: Constructive Combinatorics.
Stillwell: Elements of Algebra: Geometry, Numbers, Equations.
Stillwell: Mathematics and Its History.
Stillwell: Numbers and Geometry. *Readings in Mathematics.*
Strayer: Linear Programming and Its Applications.
Thorpe: Elementary Topics in Differential Geometry.
Toth: Glimpses of Algebra and Geometry.

Troutman: Variational Calculus and Optimal Control. Second edition.
Valenza: Linear Algebra: An Introduction to Abstract Mathematics.
Whyburn/Duda: Dynamic Topology.
Wilson: Much Ado About Calculus.