The Foundations of Geometry
Second Edition

Gerard A. Venema

Department of Mathematics and Statistics
Calvin College

PEARSON
Prentice Hall

Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data
Venema, Gerard.
The foundations of geometry / Gerard A. Venema. -- 2nd ed.
p. cm.
Includes bibliographical references index.
ISBN 0-13-602058-5
1. Geometry--Foundations--Textbooks. I. Title.

QA445.V455 2012 2012
514--dc22 2011002700 CIP

Acquisitions Editor: *Marnie Greenhut*
Assistant Editor: *Elle Driska*
Senior Managing Editor: *Karen Wernholm*
Production Project Manager: *Beth Houston*
Marketing Manager: *Roxanne McCarley*
Marketing Assistant: *Caitlin Crane*
Senior Author Support/Technology Specialist: *Joe Vetere*
Senior Manufacturing Buyer: *Carol Melville*
Design Manager: *Andrea Nix*
Senior Design Specialist: *Heather Scott*
Cover Image Credit: *Science and Society/Superstock*

Pearson Prentice Hall
Pearson Education, Inc.
Upper Saddle River, New Jersey 07458

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-602058-5

Pearson Education LTD., *London*
Pearson Education Australia PTY, Limited, *Sydney*
Pearson Education Singapore, Pte. Ltd
Pearson Education North Asia Ltd, *Hong Kong*
Pearson Education Canada, Ltd., *Toronto*
Pearson Educacion de Mexico, S.A. de C.V.
Pearson Education - Japan, *Tokyo*
Pearson Education Malaysia, Pte. Ltd

To my family:

Patricia, Sara, Emily, Daniel, David, Christian, and Ian.

Contents

APPENDICES

Preface

This is a textbook for an undergraduate course in axiomatic geometry. The text is targeted at mathematics students who have completed the calculus sequence and perhaps a first course in linear algebra, but who have not yet encountered such upper-level mathematics courses as real analysis and abstract algebra. A course based on this book will enrich the education of all mathematics majors and will ease their transition into more advanced mathematics courses. The book also includes emphases that make it especially appropriate as the textbook for a geometry course taken by future high school mathematics teachers.

WHAT'S NEW IN THE SECOND EDITION?

For the benefit of those who have used the first edition of this book, here is a quick summary of what has changed in the second edition.

- The first part of the text has been extensively reorganized and streamlined to make it possible to reach the chapter on neutral geometry more quickly.
- More exercises have been added throughout.
- The technology sections have been rewritten to facilitate the use of GeoGebra.
- The review of proof writing has been incorporated into the chapter on axiomatic systems.
- Many of the theorems in the early chapters are first stated in an informal, intuitive way and then are formally restated in if-then form.
- The organization of the chapter on neutral geometry has been tightened up.
- The coverage of asymptotically parallel lines in hyperbolic geometry was expanded and there is now a complete proof of the classification of parallels.
- A section on similarity transformations was added.
- The material on set theory and the real numbers was moved to an appendix.
- The description of various axiom systems for elementary geometry that came at the beginning of the chapter on Axioms for Plane Geometry has been expanded and moved to a new appendix where it can be covered at any time during the course.

THE FOUNDATIONS OF GEOMETRY

The principal goal of the text is to study the foundations of geometry. That means returning to the beginnings of geometry, exposing exactly what is assumed there, and building the entire subject on those foundations. Such careful attention to the foundations has a long tradition in geometry, going back more than two thousand years to Euclid and the ancient Greeks. Over the years since Euclid wrote his famous *Elements*, there have been profound changes in the way in which the foundations have been understood. Most of those changes have been byproducts of efforts to understand the true place of Euclid's parallel postulate in the foundations, so the parallel postulate is one of the primary emphases of this book.

ORGANIZATION OF THE BOOK

The book begins with a brief look at Euclid's *Elements*, and Euclid's method of organization is used as motivation for the concept of an axiomatic system. A system of axioms for geometry is then carefully laid out. The axioms used here are based on the real numbers, in the spirit of Birkhoff, and their statements have been kept as close to those in contemporary high school textbooks as is possible.

After the axioms have been stated and certain foundational issues faced, neutral geometry, in which no parallel postulate is assumed, is extensively explored. Next both Euclidean and hyperbolic geometries are investigated from an axiomatic point of view. In order to get as quickly as possible to some of the interesting results of non-Euclidean geometry, the first part of the book focuses exclusively on results regarding lines, parallelism, and triangles. Only after those topics have been treated separately in neutral, Euclidean, and hyperbolic geometries are results on area, circles, and construction introduced. While the treatment of these subjects does not exactly follow Euclid, it roughly parallels Euclid in the sense that Euclid collected most of his propositions about area in Book II and most of his propositions about circles in Books III and IV. The three chapters covering area, circles, and construction complete the coverage of the major theorems of Books I through VI of the *Elements*.

The more modern notion of a transformation is introduced next and some of the standard results regarding transformations of the plane are explored. A complete proof of the classification of the rigid motions of both the Euclidean and hyperbolic planes is included. There is a discussion of how the foundations of geometry can be reorganized to reflect the transformational point of view (as is common practice in contemporary high school geometry textbooks). Specifically, it is possible to replace the Side-Angle-Side Postulate with a postulate that asserts the existence of certain reflections.

The standard models for hyperbolic geometry are carefully constructed and the results of the chapter on transformations are used to verify their properties. The chapter on models can be relatively short because all the hard technical work involved in the constructions is done in the preceding chapter. The final chapter includes a study of some of the polygonal models that have recently been developed to help students understand what it means to say that hyperbolic space is negatively curved. The book ends with a discussion of the practical significance of non-Euclidean geometry and a brief look at the geometry of the real world.

NATIONAL STANDARDS

A significant portion of the audience for a course based on this text consists of future high school geometry teachers. In order to meet the needs of that group of mathematics majors, current national standards regarding the mathematical education of teachers and the content of the high school geometry curriculum were consulted in the design of the text. An important secondary goal of the text is to implement the recommendations in two recent sets of standards: the 2001 report on *The Mathematical Education of Teachers* [6] and the 2010 *Common Core State Standards for Mathematics* [7].

The recommendations of *The Mathematical Education of Teachers* (MET) are based on the "Principles and Standards for School Mathematics" of the National Council of Teachers of Mathematics [36]. The principal recommendation of MET is that "Prospective teachers need mathematics courses that develop a deep understanding of the mathematics they will teach" [6, Part I, page 7]. This text is designed to do precisely that in the area of geometry. A second basic recommendation in MET is that courses for prospective mathematics teachers should make explicit connections with high school mathematics. Again, this book attempts to implement that recommendation in geometry. The goal is to follow the basic recommendations of MET, not necessarily to cover every geometric topic that future teachers need to see; some geometric topics will be included in other courses, such as linear algebra.

An example of the way in which connections with high school geometry have influenced the design of the text is the choice of the axioms that are used as the starting point. The axioms on which the development of the geometry in the text is based are almost exactly those that are used in high school textbooks. While most high school textbooks still include an axiomatic treatment of geometry, there is no standard set of axioms that is common to all high school geometry courses. Therefore, various axiom systems are considered in an appendix and the merits and advantages of each are discussed. The axioms on which this text is ultimately based are as close as possible to those in contemporary high school textbooks. One of the main goals of the text is to help preservice teachers understand the logical foundations of the geometry course they will teach and that can best be accomplished in the context of axioms that are like the ones they will encounter later in the classroom. There are many other connections with high school geometry that are brought in as the text progresses.

The *Common Core State Standards for Mathematics* (CCSSM) specify what should be included in the high school geometry curriculum. This book attempts to give future teachers a grounding in the themes and perspectives described there. In particular, there is an emphasis on Euclidean geometry and the parallel postulate. The transformational approach to congruence and similarity, the approach that is promoted by CCSSM, is studied in Chapter 10 and is related to other, more traditional, ways of interpreting congruence. All of the specific topics listed in CCSSM are covered in the text. Finally, CCSSM mentions that "...in college some students will develop Euclidean and other geometries carefully from a small set of axioms." A course based on this textbook is exactly the kind of course envisioned in that remark.

Some of the newer high school mathematics curricula present mathematics in an integrated way that emphasizes connections between the branches of mathematics. There is no separate course in geometry, but rather a geometry thread is woven into all the high school mathematics courses. In order to teach such a course well, the teacher herself needs to have an understanding of the structure of geometry as a coherent subject. This book is intended to provide such an understanding.

One of the recurring themes in MET is the recommendation that prospective teachers must acquire an understanding of high school mathematics that goes well beyond that of a typical high school graduate. One way in which such understanding of geometry is often measured is in terms of the van Hiele model of geometric thought. This model is described in Appendix D. The goal of most high school courses is to develop student thinking to Level 3. A goal of this text is to bring students to Level 4 (or to Level 5, depending on whether the first level is numbered 0 or 1). It is recognized, however, that not all students entering the course are already at Level 3 and so the early part of the text is designed to ensure that students are brought to that level first.

PROOFS

A third goal of the text is to teach the art of writing proofs. There is a growing recognition of the need for a course in which mathematics students learn how to write good proofs. Such a course should serve as a bridge between the lower-level mathematics courses, which are largely technique oriented, and the upper-level courses, which tend to be much more conceptual. This book uses geometry as the vehicle for helping students to write and appreciate proofs. The ability to write proofs is a skill that can only be acquired by actually practicing it, so most of the material on writing proofs is integrated into the text and the attention to proof permeates the entire text. This means that the book can also be used in classes where the students already have experience writing proofs; despite the emphasis on writing proofs, the book is still primarily a geometry text.

Having the geometry course serve as the introduction to proof represents a return to tradition in that the course in Euclidean geometry has for thousands of years been seen as the standard introduction to logic, rigor, and proof in mathematics. Using the geometry course this way makes historical sense because the axiomatic method was first introduced in geometry and geometry remains the branch of mathematics in which that method has had its greatest success. While proof and logical deduction are still emphasized in the standards for high school mathematics, most high school students no longer take a full-year course devoted exclusively to geometry with a sustained emphasis on proof. This makes it more important than ever that we teach a good college-level geometry course to all mathematics students. By doing so we can return geometry to its place as the subject in which students first learn to appreciate the importance of clearly spelling out assumptions and deducing results from those assumptions via careful logical reasoning.

The emphasis on proof makes the course a do-it-yourself course in that the reader will be asked to supply proofs for many of the key theorems. Students who diligently work the exercises come away from the course with a sense that they have an unusually deep understanding of the material. In this way the student should not only learn the mechanics of good proof writing style but should also come to more fully appreciate the important role proof plays in an understanding of mathematics.

HISTORICAL AND PHILOSOPHICAL PERSPECTIVE

A final goal of the text is to present a historical perspective on geometry. Geometry is a dynamic subject that has changed over time. It is a part of human culture that was created and developed by people who were very much products of their time and place. The foundations of geometry have been challenged and reformulated over the years, and beliefs about the relationship between geometry and the real world have been challenged as well.

The material in the book is presented in a way that is sensitive to such historical and philosophical issues. This does not mean that the material is presented in a strictly historical order or that there are lengthy historical discussions but rather that geometry is presented in such a way that the reader can understand and appreciate the historical development of the subject and so that it would be natural to investigate the history of the subject while learning it. Many chapters include suggested readings on the history of geometry that can be used to enrich the text.

Throughout the book there are references to philosophical issues that arise in geometry. For example, one question that naturally occurs to anyone studying non-Euclidean geometry is this: What is the connection between the abstract entities that are studied in a course on the foundations of geometry and properties of physical space? The book does not present dogmatic answers to such questions, but instead simply raises them in an effort to promote student thinking. The hope is that this will serve to counter the common perception that mathematics is a subject in which every question has a single correct answer and in which there is no room for creative ideas or opinions.

WHY AXIOMATIC GEOMETRY?

One question I am often asked is: Why study axiomatic geometry? Why take such an old-fashioned approach to geometry when there are so many beautiful and exciting modern topics that could be included in the course? The main answer I give is that proof and the axiomatic method remain hallmarks of mathematics. In order to be well educated in mathematics, students should see a full axiomatic development of a complete branch of mathematics. They need to know about the historical importance of the axiomatic approach in geometry and they need to be aware of the profound changes in our understanding of the relationship between mathematics and the real world that grew out of attempts to understand the place of the parallel postulate.

For many reasons, a course in axiomatic (Euclidean) geometry is a natural setting in which students can learn to write and appreciate proofs.

- The objects studied are natural and familiar.
- The definitions are uncomplicated, requiring few quantifiers.
- All necessary assumptions can be completely described.
- The proofs are relatively straightforward and the ideas can be understood visually.
- The proofs usually require a nontrivial idea, so students appreciate the need for a proof.

The proofs in the early books of Euclid's *Elements* are beautiful, yet simple, containing just the right amount of detail. Studying the proofs of Euclid remains one of the best introductions to proofs and the geometry course can serve as a bridge to higher mathematics for all mathematics majors. Another justification for a college course in axiomatic geometry is that most students no longer have the experience of studying axiomatic geometry in high school.

WHY EUCLIDEAN GEOMETRY?

Another question I am frequently asked is: Why include such an extensive treatment of Euclidean geometry in a college course? Don't students learn enough about that subject in high school? The answer, sadly, is that students are not learning enough about Euclidean geometry in high school. Most high school geometry courses no longer include a study of the axioms and some do not emphasize proof at all. In many cases there is not even a separate course in geometry, but geometry is one of several threads that are woven together in the high school mathematics curriculum. Rather than lamenting these developments, I think college and university mathematics departments should embrace the opportunity to teach a substantial geometry course at the college level. This can restore Euclidean geometry to its traditional role as the course in which mathematics students have their first experience with careful logical thinking and the complete development of a comprehensive, coherent mathematical subject.

TECHNOLOGY

In recent years powerful computer software has been developed that can be used to explore geometry. The study of geometry from this book can be greatly enhanced by such dynamic software and the reader is encouraged to find appropriate ways in which to incorporate this technology into the geometry course. While software can enrich the experience of learning geometry from this book, its use is not required. The book can be read and studied quite profitably without it.

The author recommends the use of the dynamic mathematics software program GeoGebra. GeoGebra is free software that is intended to be used for teaching and learning mathematics. It may be downloaded from the website www.geogebra.org. The software has many great features that make it ideal for use in the geometry classroom, but the main advantage it has over commercial geometry software is the fact that it is free and runs under any of the standard computer operating systems. This means that students can load the program on their own computers and will always have access to it.

Any of the several commercially available pieces of dynamic geometry software will also serve the purpose. *Geometer's Sketchpad*™ (Key Curriculum Press) is widely used and readily available. *Cabri Geometry*™ (Texas Instruments) is less commonly used in college-level courses, but it is also completely adequate. It has some predefined tools, such as an inversion tool and a test for collinearity, that are not included in Sketchpad. *Cinderella*™ (Springer-Verlag) is Java-based software and is also very good. It has the advantage that it allows diagrams to be drawn in all three two-dimensional geometries: Euclidean, hyperbolic, and spherical. Another advantage is that it allows diagrams to be easily exported as Java applets. A program called NonEuclid is freely available on the internet and it can be used to enhance the non-Euclidean geometry in the course. New software is being produced all the time, so you may find that other products are available to you.

This is a course in the foundations of axiomatic geometry, and software will necessarily play a more limited role in such a course than it might in other kinds of geometry courses. Nonetheless, there is an appropriate role for software in a course such as this and the author hopes that the book will demonstrate that. There is no reason for those who love the proofs of Euclid to resist the use of technology. After all, Euclid himself made use of the limited technology that was available to him, namely the compass and straightedge. In the same way we can make good use of modern technology in our study of geometry. It is especially important that future high school teachers learn to understand and appreciate the *appropriate* use of technology.

In the first part of the text (Chapters 2 through 4), the objective is to carefully expose all the assumptions that form the foundations of geometry and to understand for ourselves how the basic results of geometry are built on those foundations. For most users the software is a black box in the sense that we either don't know what assumptions are built into it or we have only the authors' description of what went into the software. As a result, software is of limited use in this part of the course and it will not be mentioned explicitly in the first four chapters of the book. But you should be using it to draw diagrams and to experiment with what happens when you vary the data in the theorems. During that phase of the course the main function of the software is to illustrate one possible interpretation of the relationships being studied.

It is in the second half of the course that the software comes into its own. Computer software is ideal for experimenting, exploring, and discovering new relationships. In order to illustrate that, several of the later chapters include sections in which the software is used to explore ideas that go beyond those that are presented in detail and to discover new relationships. In particular, there are such exploratory sections in the chapters on Euclidean geometry and circles. The entire chapter on constructions is written as an exploration with only a limited number of proofs or hints provided in the text.

The exploratory sections of the text have been expanded into a laboratory manual entitled *Exploring Advanced Euclidean Geometry with GeoGebra*.

SUPPLEMENTS

Two supplements are available: an *Instructors' Manual* and a computer laboratory manual entitled *Exploring Advanced Euclidean Geometry with GeoGebra*.

- The *Instructors' Manual* contains complete solutions to all of the exercises as well as information about how to teach from the book. Instructors may download it from
  http://www.pearsonhighered.com/irc.

- Additional materials may be downloaded, free of charge, from the author's website at
  http://calvin.edu/~venema/geometrybook.html.

DESIGNING A COURSE

A full-year course should cover essentially all the material in the text. There can be some variation based on instructor and student interest, but most or all of every chapter should be included.

An instructor teaching a one-semester or one-quarter course will be forced to pick and choose. It is important that this be done carefully so that the course reaches some of the interesting and useful material that is to be found in the second half of the book.

- Chapter 1 sets the stage for what is to come, so it should be covered in some way. But it can be discussed briefly in class and then assigned as reading.
- Chapter 2 should definitely be covered because it establishes the basic framework for the treatment of geometry that follows.
- The basic coverage of geometry begins with Chapter 3. Chapters 3 and 4 form the heart of a one-semester course. Those chapters should be included in any course taught from the book.
- At least some of Chapter 5 should also be included in any course.
- Starting with Chapter 7, the chapters are largely independent of each other and an instructor can select material from them based on the interests and needs of the class.

Several sample course outlines are included below. Many other variations are possible. It should be noted that the suggested outlines are ambitious and many instructors will choose to cover less.

A course emphasizing Euclidean Geometry

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	$\leqslant 2$
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	2
7	Area	1–2
8	Circles	1–2
10	Transformations	2

A course emphasizing non-Euclidean Geometry

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	$\leqslant 2$
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	1
6	Hyperbolic geometry	2
7	Area	1–2
11	Models	1–2
12	Geometry of space	1

A course for future high school teachers

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	$\leqslant 2$
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	1
6	Hyperbolic geometry	1
7	Area	1
8	Circles	1
10	Transformations	1
11	Models	1
12	Geometry of space	1

The suggested course for future high school teachers includes just a brief introduction to each of the topics in later chapters. The idea is that the course should provide enough background so that students can study those topics in more depth later if they need to. It is hoped that this book can serve as a valuable reference for those who go on to teach geometry courses. The book could be a resource that provides information about rigorous treatments of such topics as parallel lines, area, circles, constructions, transformations, and so on, that are part of the high school curriculum.

ACKNOWLEDGMENTS

I would like to thank all those who helped me put this book together. I should again recognize the individuals who were acknowledged in the Preface to the first edition: Ric Ancel, Ginni Baker, Lawrence Cannon, Tom Clark, Manfred Dugas, Patricia Hale, Adam Lewenberg, George Lobell, Mario Lopez, Todd Moyer, F. Alexander Norman, David Royster, Franz Rothe, Peter F. Stiller, Mike Stob, and Anke Walz. Since that edition appeared, many people who were using the book have contacted me and offered helpful comments. All of the comments, questions, and suggestions from instructors and students alike are appreciated and have contributed to this second edition. Unfortunately there were too many of them for me to list them individually.

I am grateful to the members of the Calvin College Foundations of Geometry classes who have allowed me to try out various versions of the text on them; the feedback they provided was invaluable.

I wish to thank the following reviewers of the second edition:

Mark Barsamian *Ohio University Athens*	Carol Seaman *UNC Greensboro*	Richard Stout *Gordon College*
Jason Johnson *Middle Tennessee State University*	Glenn Sproul *Johnson State College*	Violeta Vasilevska *Utah Valley University*
David Royster *University of Kentucky*	David Smith *University of Texas Arlington*	Stephen Wilson *Johns Hopkins University*

Finally, thanks to Sean Lawton and Ethan Pribble for their work as accuracy checkers, as well as Marnie Greenhut, Elle Driska, and all the members of the staff at Pearson who helped make this book possible.

Gerard A. Venema
Calvin College
May 2011

CHAPTER 1

Prologue: Euclid's *Elements*

Our study of geometry begins with an examination of the historical origins of the axiomatic method in geometry. While the material in this chapter is not a mathematical prerequisite for what comes later, an appreciation of the historical roots of axiomatic thinking is essential to an understanding of why the foundations of geometry are systematized as they are.

1.1 GEOMETRY BEFORE EUCLID

Geometry is an ancient subject. Its roots go back thousands of years and geometric ideas of one kind or another are found in nearly every human culture. The beauty of geometric patterns is universally appreciated and often investigated in a systematic way. The study of geometry as we know it emerged more than 4000 years ago in Mesopotamia, Egypt, India, and China.

Because the Nile River annually flooded vast areas of land and obliterated property lines, surveying and measuring were important to the ancient Egyptians. This practical interest may have motivated their study of geometry. Egyptian geometry was mostly an empirical science, consisting of many rule-of-thumb procedures that were arrived at through experimentation, observation, and trial and error. Formulas were approximate ones that appeared to work, or at least gave answers that were close enough for practical purposes. But the ancient Egyptians were also aware of more general principles, such as special cases of the Pythagorean Theorem and formulas for volumes.

The ancient Mesopotamians, or Babylonians, had an even more advanced understanding of geometry. They knew the Pythagorean Theorem long before Pythagoras. They discovered some of the area-based proofs of the theorem that will be discussed in Chapter 7, and knew a general method that generates all triples of integers that are lengths of sides of right triangles. In India, ancient texts apply the Pythagorean Theorem to geometric problems associated with the design of structures. The Pythagorean Theorem was also discovered in China at roughly the same time.

About 2500 years ago there was a profound change in the way geometry was practiced: Greek mathematicians introduced abstraction, logical deduction, and proof into geometry. They insisted that geometric results be based on logical reasoning from first principles. In theory this made the results of geometry exact, certain, and undeniable, rather than just likely or approximate. It also took geometry out of the realm of everyday experience and made it a subject that studies abstract entities. Since the purpose of this course is to study the logical foundations of geometry, it is natural that we should start with the geometry of the ancient Greeks.

The process of introducing logic into geometry apparently began with Thales of Miletus around 600 B.C. and culminated in the work of Euclid of Alexandria in approximately 300 B.C. Euclid is the most famous of the Greek geometers and his name is still universally associated with the geometry that is studied in schools today. Most of the ideas that are included in what we call "Euclidean Geometry" probably did not originate with Euclid himself; rather, Euclid's contribution was to organize and present the results of Greek geometry in a logical and coherent way. He published his results in a series of thirteen books known as his *Elements*. We begin our study of geometry by examining those *Elements* because they set the agenda for geometry for the next two millennia and more.

1.2 THE LOGICAL STRUCTURE OF EUCLID'S *ELEMENTS*

Euclid's *Elements* are organized according to strict logical rules. Euclid begins each book with a list of definitions of the technical terms he will use in that book. In Book I he next states five "postulates" and five "common notions." These are assumptions that are meant to be accepted without proof. Both the postulates and common notions are basic statements whose truth should be evident to any reasonable person. They are the starting point for what follows. Euclid recognized that it is not possible to prove everything, that he had to start somewhere, and he attempted to be clear about exactly what his assumptions were.

Most of Euclid's postulates are simple statements of intuitively obvious and undeniable facts about space. For example, Postulate I asserts that it is possible to draw a straight line through any two given points. Postulate II says that a straight line segment can be extended to a longer segment. Postulate III states that it is possible to construct a circle with any given center and radius. Traditionally these first three postulates have been associated with the tools that are used to implement them on a piece of paper. The first two postulates allow two different uses of a straight edge: A straight edge can be used to draw a line segment connecting any two points or to extend a given line segment to a longer one. The third postulate affirms that a compass can be used to construct a circle with a given center and radius. Thus the first three postulates simply permit the familiar straightedge and compass constructions of high school geometry.

The fourth postulate asserts that all right angles are congruent ("equal" in Euclid's terminology). The fifth postulate makes a more subtle and complicated assertion about two lines that are cut by a transversal. These last two postulates are the two technical facts about geometry that Euclid needs in his proofs.

The common notions are also intuitively obvious facts that Euclid plans to use in his development of geometry. The difference between the common notions and the postulates is that the common notions are not peculiar to geometry but are common to all branches of mathematics. They are everyday, common-sense assumptions. Most spell out properties of equality, at least as Euclid used the term *equal*.

The largest part of each Book of the *Elements* consists of propositions and proofs. These too are organized in a strict, logical progression. The first proposition is proved using only the postulates, Proposition 2 is proved using only the postulates and Proposition 1, and so on. Thus the entire edifice is built on just the postulates and common notions; once these are granted, everything else follows logically and inevitably from them. What is astonishing is the number and variety of propositions that can be deduced from so few assumptions.

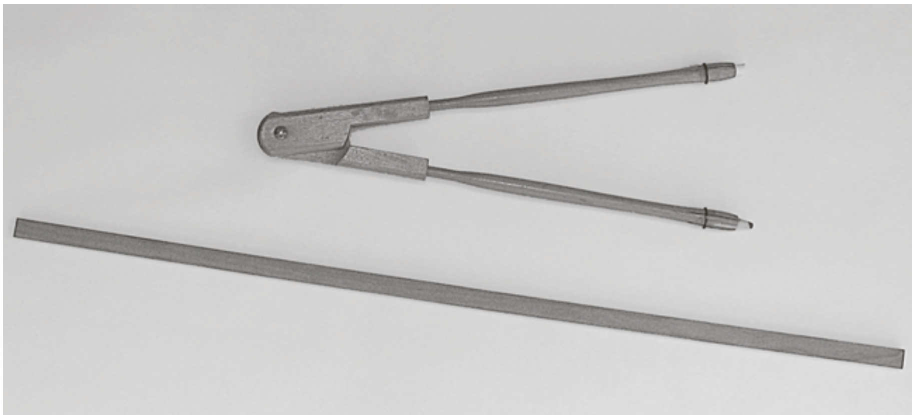

FIGURE 1.1: Euclid's tools: a compass and a straightedge

1.3 THE HISTORICAL SIGNIFICANCE OF EUCLID'S *ELEMENTS*

It is nearly impossible to overstate the importance of Euclid's *Elements* in the development of mathematics and human culture generally. From the time they were written, the *Elements* have been held up as the standard for the way in which careful thought ought to be organized. They became the model for the development of all scientific and philosophical theories. What was especially admired about Euclid's work was the way in which he clearly laid out his assumptions and then used pure logic to deduce an incredibly varied and extensive set of conclusions from them.

Up until the twentieth century, Euclid's *Elements* were the textbook from which all students learned both geometry and logic. Even today the geometry in school textbooks is presented in a way that is remarkably close to that of Euclid. Furthermore, much of what mathematicians did during the next two thousand years centered around tying up loose ends left by Euclid. Countless mathematicians spent their careers trying to solve problems that were raised by Greek geometers of antiquity and trying to improve on Euclid's treatment of the foundations.

Most of the efforts at improvement focused on Euclid's fifth postulate. Even though the statement does not explicitly mention parallel lines, this postulate is usually referred to as "Euclid's parallel postulate." It asserts that two lines that are cut by a transversal must intersect on one side of the transversal if the interior angles on that side of the transversal sum to less than two right angles. In particular, it proclaims that the condition on the angles formed by a transversal implies that the two given lines are not parallel. Thus it is really a statement about nonparallel lines. As we shall see later, the postulate can be reformulated in ways that make it more obviously and directly a statement about parallel lines.

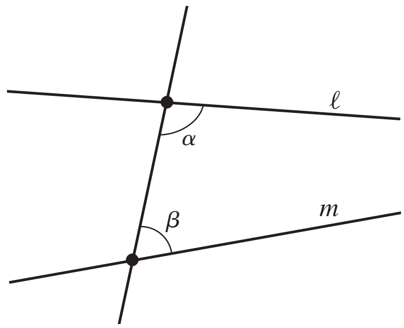

FIGURE 1.2: Euclid's Postulate V. If the sum of α and β is less than two right angles, then ℓ and m must eventually intersect

A quick reading of the postulates (see the next section) reveals that Postulate V is noticeably different from the others. For one thing, its statement is much longer than those of the other postulates. A more significant difference is the fact that it involves a fairly complicated arrangement of lines and also a certain amount of ambiguity in that the lines must be "produced indefinitely." It is not as intuitively obvious or self-evident as the other postulates; it has the look and feel of a proposition rather than a postulate. For these reasons generations of mathematicians tried to improve on Euclid by attempting to prove that Postulate V is a logical consequence of the other postulates or, failing that, they tried at least to replace Postulate V with a simpler, more intuitively obvious postulate from which Postulate V could then be deduced as a consequence. No one ever succeeded in proving the fifth postulate using just the first four postulates, but it was not until the nineteenth century that mathematicians fully understood why that was the case.

It should be recognized that these efforts at improvement were not motivated by a perception that there was anything wrong with Euclid's work. Quite the opposite: Thousands of mathematicians spent enormous amounts of time trying to improve on Euclid precisely because they thought so highly of Euclid's accomplishments. They wanted to take what was universally regarded as the crown of theoretical thought and make it even more wonderful than it already was!

Another important point is that efforts to rework Euclid's treatment of geometry led indirectly to progress in mathematics that went far beyond mere improvements in the *Elements* themselves. Attempts to prove Euclid's Fifth Postulate eventually resulted in the realization that, in some kind of stroke of genius, Euclid somehow had the great insight to pinpoint one of the deepest properties that a geometry may have. Not only that, but it was discovered that there are alternative geometries in which Euclid's Fifth Postulate fails to hold. These discoveries were made in the early nineteenth century and had far-reaching implications for all of mathematics. They opened up whole new fields of mathematical study; they also produced a revolution in the conventional view of how mathematics relates to the real world and forced a new understanding of the nature of mathematical truth.

The story of how Euclid's Parallel Postulate inspired all these developments is one of the most interesting in the history of mathematics. That story will unfold in the course of our study of geometry in this book. It is only in the light of that story that the current organization of the foundations of geometry can be properly understood.

1.4 A LOOK AT BOOK I OF THE *ELEMENTS*

In order to give more substance to our discussion, we now take a direct look at parts of Book I of the *Elements*. All Euclid's postulates are stated below as well as selected definitions and propositions. The excerpts included here are chosen to illustrate the points that will be made in the following section. The translation into English is by Sir Thomas Little Heath (1861–1940).

Some of Euclid's Definitions

Definition 1. A *point* is that which has no part.

Definition 2. A *line* is breadthless length.

Definition 4. A *straight line* is a line which lies evenly with the points on itself.

Definition 10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is *right*, and the straight line standing on the other is called a *perpendicular* to that on which it stands.

Definition 11. An *obtuse* angle is an angle greater than a right angle.

Definition 12. An *acute* angle is an angle less than a right angle.

Euclid's Postulates

Postulate I. *To draw a straight line from any point to any point.*

Postulate II. *To produce a finite straight line continuously in a straight line.*

Postulate III. *To describe a circle with any center and distance.*

Postulate IV. *That all right angles are equal to one another.*

Postulate V. *That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.*

Euclid's Common Notions

Common Notion I. *Things which equal the same thing are also equal to one another.*

Common Notion II. *If equals be added to equals, the wholes are equal.*

Common Notion III. *If equals be subtracted from equals, the remainders are equal.*

Common Notion IV. *Things which coincide with one another are equal to one another.*

Common Notion V. *The whole is greater than the part.*

Three of Euclid's Propositions and their proofs

Proposition 1. *On a given finite straight line to construct an equilateral triangle.*

Let AB be the given finite straight line. Thus it is required to construct an equilateral triangle on the straight line AB. With center A and distance AB let the circle BCD be described [Post. III]; again, with center B and distance BA let the circle ACE be described [Post. III]; and from the point C, in which the circles cut one another, to the points A, B let the straight lines CA, CB be joined [Post. I].

Now, since the point A is the center of the circle CDB, AC is equal to AB [Def. 15]. Again, since the point B is the center of the circle CAE, BC is equal to BA [Def. 15]. But CA was also proved equal to AB; therefore each of the straight lines CA, CB is equal to AB. And things which are equal to the same thing also equal one another [C.N. I]; therefore CA is also equal to CB. Therefore the three straight lines CA, AB, BC are equal to one another. Therefore the triangle ABC is equilateral; and it has been constructed on the given finite straight line AB.

Being what it was required to do.

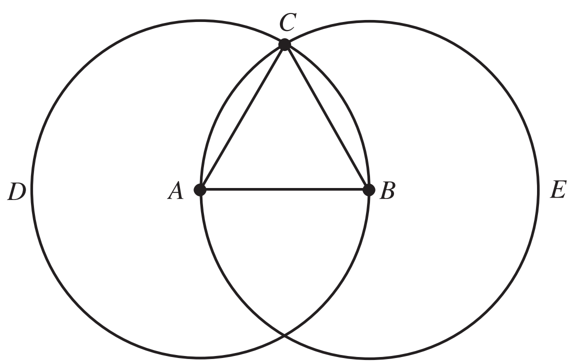

FIGURE 1.3: Euclid's diagram for Proposition 1

Proposition 4. *If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.*

Let ABC, DEF be two triangles having the two sides AB, AC equal to the two sides DE, DF respectively, namely AB to DE and AC to DF, and the angle BAC equal to the angle EDF. I say that the base BC is also equal to the base EF, the triangle ABC will be equal to the triangle DEF, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend, that is, the angle ABC to the angle DEF, and the angle ACB to the angle DFE.

For, if the triangle ABC be applied to the triangle DEF, and if the point A be placed on the point D and the straight line AB on DE, then the point B will also coincide with E, because AB is equal to DE. Again, AB coinciding with DE, the straight line AC will also coincide with DF, because the angle BAC is equal to the angle EDF; hence the point C will also coincide with the point F, because AC is again equal to DF.

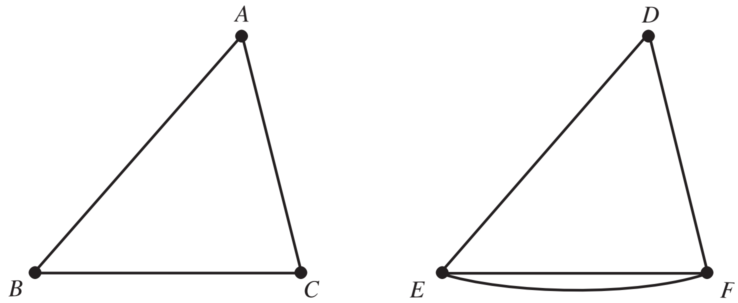

FIGURE 1.4: Euclid's diagram for Proposition 4

But B also coincided with E; hence the base BC will coincide with the base EF. **[**For if, when B coincides with E and C with F, the base BC does not coincide with the base EF, two straight lines will enclose a space: which is impossible. Therefore the base BC will coincide with EF**]** and will be equal to it [C.N. 4]. Thus the whole triangle ABC will coincide with the whole triangle DEF, and will be equal to it. And the remaining angles also coincide with the remaining angles and will be equal to them, the angle ABC to the angle DEF, and the angle ACB to the angle DFE.

Therefore etc. Being what it was required to prove.

Proposition 16. *In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.*

Let ABC be a triangle, and let one side of it BC be produced to D; I say that the exterior angle ACD is greater than either of the interior and opposite angles CBA, BAC.

Let AC be bisected at E [Prop. 10] and let BE be joined and produced in a straight line to F; let EF be made equal to BE [Prop. 3], let FC be joined [Post. I], and let AC be drawn through to G [Post. II]. Then since AE is equal to EC, and BE to EF, the two sides AE, EB are equal to the two sides CE, EF respectively; and the angle AEB is equal to the angle FEC, for they are vertical angles [Prop. 15]. Therefore the base AB is equal to the base FC, and the triangle ABE is equal to the triangle CFE, and the remaining angles are equal to the remaining angles respectively, namely those which the equal sides subtend [Prop. 4]; therefore the angle BAE is equal to the angle ECF. But the angle ECD

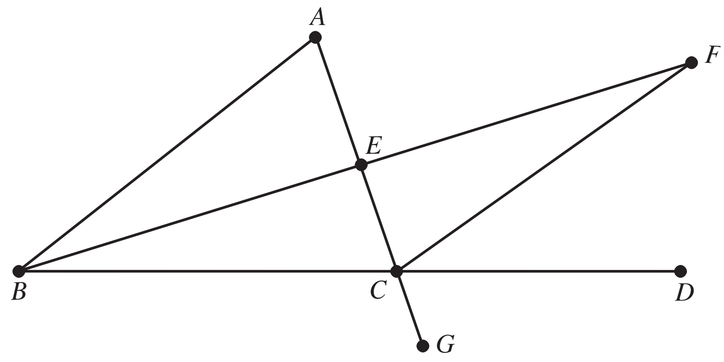

FIGURE 1.5: Euclid's diagram for Proposition 16

is greater than the angle ECF [C.N. 5]; therefore the angle ACD is greater than the angle BAE.

Similarly, also if BC is bisected, the angle BCG, that is, the angle ACD [Prop. 15], can be proved greater than the angle ABC as well.

Therefore etc. Q.E.D.

1.5 A CRITIQUE OF EUCLID'S *ELEMENTS*

As indicated earlier, the *Elements* have been the subject of a great deal of interest over the thousands of years since Euclid wrote them and the study of Euclid's method of organizing his material has inspired mathematicians to even greater levels of logical rigor. Originally attention was focused on Euclid's postulates, especially his fifth postulate, but efforts to clarify the role of the postulates eventually led to the realization that there are difficulties with other parts of Euclid's *Elements* as well. When the definitions and propositions are examined in the light of modern standards of rigor it becomes apparent that Euclid did not achieve all the goals he set for himself—or at least that he did not accomplish everything he was traditionally credited with having done.

Euclid purports to define all the technical terms he will use.[1] However, an examination of his definitions shows that he did not really accomplish this. The first few definitions are somewhat vague, but suggestive of intuitive concepts. An example is the very first definition, in which *point* is defined as "that which has no part." This does indeed suggest something to most people, but it is not really a rigorous definition in that it does not stipulate what sorts of objects are being considered. It is somehow understood from the context that it is only geometric objects which cannot be subdivided that are to be called points. Even then it is not completely clear what a point is: Apparently a point is pure location and has no size whatsoever. But there is nothing in the physical world of our experience that has those properties exactly. Thus we must take point to be some kind of idealized abstract entity and admit that its exact nature is not adequately explained by the definition. Similar comments could be made about Euclid's definitions of *line* and *straight line*.

By contrast, later definitions are more complete in that they define one technical word in terms of others that have been defined previously. Examples are Definitions 11 and 12 in which *obtuse angle* and *acute angle* are defined in terms of the previously defined

[1] Some scholars suggest that the definitions included in the versions of the *Elements* that have come down to us were not in Euclid's original writings, but were added later. Even if that is the case, the observations made here about the definitions are still valid.

right angle. From the point of view of modern rigor there is still a gap in these definitions because Euclid does not specify what it means for one angle to be greater than another. The difference is that these definitions are complete in themselves and would be rigorous and usable if Euclid were to first spell out what it means for one angle to be greater than another and also define what a right angle is.

Such observations have led to the realization that there are actually two kinds of technical terms. It is not really possible to define all terms; just as some statements must be accepted without proof and the other propositions proved as consequences, so some terms must be left undefined. Other technical terms can then be defined using the undefined terms and previously defined terms. This distinction will be made precise in the next chapter.

A careful reading of Euclid's proofs reveals some gaps there as well. The proof of Proposition 1 is a good example. In one sense, the proposition and its proof are simple and easy to understand. In modern terminology the proposition asserts the following: *Given two points A and B, it is possible to construct a third point C such that* $\triangle ABC$ *is an equilateral triangle.* Euclid begins with the segment from A to B. He then uses Postulate III to draw two circles of radius AB, one centered at A and the other centered at B. He takes C to be one of the two points at which the circles intersect and uses Postulate I to fill in the sides of a triangle. Euclid completes the proof by using the common notions to explain why the triangle he has constructed must be equilateral. The written proof is supplemented by a diagram that makes the construction clear and convincing.

Closer examination shows, however, that Euclid assumed more than just what he stated in the postulates. In particular there is nothing explicitly stated in the postulates that would guarantee the existence of a point C at which the two circles intersect. The existence of C is taken for granted because the diagram clearly shows the circles intersecting in two points. There are, however, situations in which no point of intersection exists; one such example will be studied in Chapter 3. So Euclid is using "facts" about his points and lines that are undoubted and intuitively obvious to most readers, but which have not been explicitly stated in the postulates.

Euclid's Proposition 4 is the familiar Side-Angle-Side Congruence Condition from high school geometry. This proposition is not just a construction like Proposition 1 but asserts a logical implication: If two sides and the included angle of one triangle are congruent to the corresponding parts of a second triangle, then the remaining parts of the two triangles must also be congruent. Euclid's method of proof is interesting. He takes one triangle and "applies" it to the other triangle. By this we understand that he means to pick up the first triangle, move it, and carefully place one vertex at a time on the corresponding vertices of the second triangle. This is often called Euclid's *method of superposition*. It is quite clear from an intuitive point of view that this operation should be possible, but again the objection can be raised that Euclid is using unstated assumptions about triangles. Over the years geometers have come to realize that the ability to move geometric objects around without distorting their shapes cannot be taken for granted. The need to include an explicit assumption about motions of triangles will be discussed further in Chapters 3 and 10.

Another interesting aspect of the proof is the fact that part of it is enclosed in square brackets. (See the words starting with, "For if ..." in the third paragraph of the proof.) These words are in brackets because it is believed that they are not part of Euclid's original proof, but were inserted later.[2] They were added to justify Euclid's obvious assumption that there is only one straight line segment joining two points. Postulate I states that there exists a straight line joining two points, but here Euclid needs the stronger statement that

[2] See [22, page 249].

there is exactly one such line. The fact that these words were added in antiquity is an indication that already then some readers of the *Elements* recognized that Euclid was using unstated assumptions.

Euclid's Proposition 16 is the result we now know as the Exterior Angle Theorem. This theorem and its proof will be discussed in Chapter 4. For now we merely point out that Euclid's proof depends on a relationship that appears to be obvious from the diagram provided, but which Euclid does not actually prove. Euclid wants to show that the interior angle $\angle BAC$ is smaller than the exterior angle $\angle ACD$. He first constructs the points E and F, and then uses the Vertical Angles Theorem (Proposition 15) and Side-Angle-Side to conclude that $\angle BAC$ is congruent to $\angle ACF$. Euclid assumes that F is in the interior of $\angle ACD$ and uses Common Notion 5 to conclude that $\angle BAC$ is smaller than $\angle ACD$. However he provides no justification for the assertion that F is in the interior of angle $\angle ACD$. In Chapter 3 we carefully state postulates that will allow us to fill in this gap when we prove the Exterior Angle Theorem in Chapter 4.

The preceding discussion is not meant to suggest that Euclid is wrong in his conclusions or that his work is in any way flawed. Rather, the point is that standards of mathematical rigor have changed since Euclid's day. Euclid thought of his postulates as statements of self-evident truths about the real world. He stated the key geometric facts as postulates, but felt free to bring in other spatial relationships when they were needed and were obvious from the diagrams. As we will see in the next chapter, we no longer see postulates in the same way Euclid did. Instead we try to use the postulates to state *all* the assumptions that are needed in order to prove our theorems. If that is our goal, we will need to assume much more than is stated in Euclid's postulates.

1.6 FINAL OBSERVATIONS ABOUT THE *ELEMENTS*

One aspect of Euclid's proofs that should be noted is the fact that each statement in the proof is justified by appeal to one of the postulates, common notions, definitions, or previous propositions. These references are placed immediately after the corresponding statements. They were probably not written explicitly in Euclid's original and therefore Heath encloses them in square brackets. This aspect of Euclid's proofs serves as an important model for the proofs we will write later in this course.

The words "Therefore etc." found near the end of the proofs are also not in Euclid's original. In the Greek view, the proof should culminate in a full statement of what had been proved. Thus Euclid's proof would have ended with a complete restatement of the conclusion of the proposition. Heath omits this reiteration of the conclusion and simply replaces it with "etc." Notice that the proof of Proposition 1 ends with the phrase "Being what it was required to do," while the proof of Proposition 4 ends with "Being what it was required to prove." The difference is that Proposition 1 is a construction while Proposition 4 is a logical implication. Later Heath uses the Latin abbreviations Q.E.F. and Q.E.D. for these phrases.

There are many features of Euclid's work that strike the modern reader as strange. One is the spare purity of Euclid's geometry. The points and lines are pure geometric forms that float in the plane with no fixed location. All of us have been trained since childhood to identify points on a line with numbers and points in the plane with pairs of numbers. That concept would have been foreign to Euclid; he did not mix the notions of number and point the way we do. The identification of number and point did not occur until the time of Descartes in the seventeenth century and it was not until the twentieth century that the real numbers were incorporated into the statements of the postulates of geometry. It is important to recognize this if we are to understand Euclid.

Euclid (really Heath) also uses language in a way that is different from contemporary usage. For example, what Euclid calls a line we would call a curve. We reserve the term *line* for what Euclid calls a "straight line." More precisely, what Euclid calls a straight line we would call a line segment (finitely long, with two endpoints). This distinction is more than just a matter of definitions; it indicates a philosophical difference. In Euclid, straight lines are potentially infinite in that they can always be extended to be as long as is needed for whatever construction is being considered, but he never considers the entire infinite line all at once. Since the time of Georg Cantor in the nineteenth century, mathematicians have been comfortable with sets that are actually infinite, so we usually think of the line as already being infinitely long and do not worry about the need to extend it.

Euclid chose to state his postulates in terms of straightedge and compass constructions. His propositions then often deal with the question of what can be constructed using those two instruments. For example, Proposition 1 really asserts the following: *Given a line segment, it is possible to construct, using only straightedge and compass, an equilateral triangle having the given segment as base.* In some ways Euclid identifies constructibility with existence. One of the major problems that the ancient Greeks never solved is the question of whether or not a general angle can be trisected. From a modern point of view the answer is obvious: any angle has a measure (in degrees, for example) which is a real number; simply dividing that real number by 3 gives us an angle that is one-third the original. But the question the Greeks were asking was whether or not the smaller angle can always be constructed from the original using only straightedge and compass. Such constructibility questions will be discussed in Chapter 9.

In this connection it is worthwhile to observe that the tools Euclid chose to use reflect the same pure simplicity that is evident throughout his work. His straightedge has no marks on it whatsoever. He did not allow a mark to be made on it that could be preserved when the straightedge is moved to some other location. In modern treatments of geometry we freely allow the use of a ruler, but we should be sure to note that a ruler is much more than a straightedge: It not only allows straight lines to be drawn, but it also measures distances at the same time. Euclid's compass, in the same way, is what we would now call a "collapsing" compass. It can be used to draw a circle with a given center and radius (where "radius" means a line segment with the center as one endpoint), but it cannot be moved to some other location and used to draw a different circle of the same radius. When the compass is picked up to be moved, it collapses and does not remember the radius of the previous circle. In contemporary treatments of geometry the compass has been supplemented by a protractor, which is a device for measuring angles. Euclid did not rely on numerical measurements of angles and he did not identify angles with the numbers that measure them the way we do.

SUGGESTED READING

1. Chapters 1 and 2 of *Journey Through Genius*, [16].
2. Part I (pages 1–49) of *Euclid's Window*, [33].
3. Chapters 1–4 of *Geometry: Our Cultural Heritage*, [26].
4. Chapters I–IV of *Mathematics in Western Culture*, [29].
5. Chapters 1 and 2 of *The Non-Euclidean Revolution*, [44].
6. Chapters 1 and 2 of *A History of Mathematics*, [27].

EXERCISES 1.6

1. A *quadrilateral* is a four-sided figure in the plane. Consider a quadrilateral whose successive sides have lengths a, b, c, and d. Ancient Egyptian geometers used the formula

$$A = \frac{1}{4}(a + c)(b + d)$$

to calculate the area of a quadrilateral. Check that this formula gives the correct answer for rectangles but not for parallelograms.

2. An ancient Egyptian document, known as the *Rhind papyrus*, suggests that the area of a circle can be determined by finding the area of a square whose side has length $\frac{8}{9}$ the diameter of the circle. What value of π is implied by this formula? How close is it to the correct value?

3. The familiar Pythagorean Theorem states that if $\triangle ABC$ is a right triangle with right angle at vertex C and a, b, and c, are the lengths of the sides opposite vertices A, B, and C, respectively, then $a^2 + b^2 = c^2$. Ancient proofs of the theorem were based on diagrams like those in Figure 1.6. Explain how the two diagrams together can be used to provide a proof for the theorem.

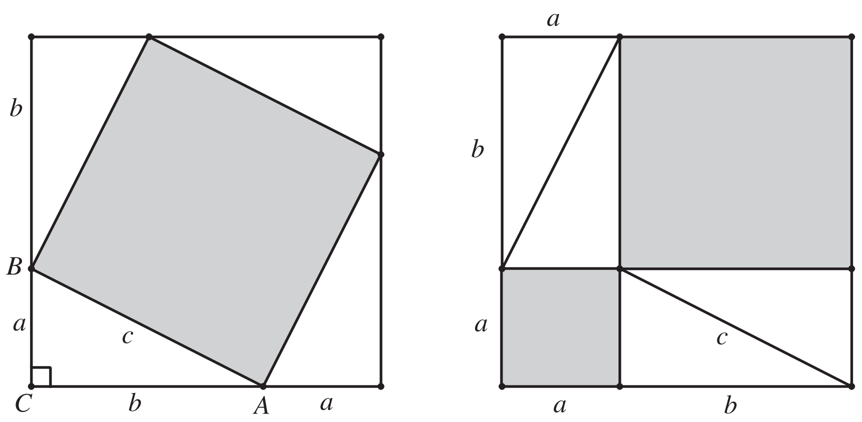

FIGURE 1.6: Proof of The Pythagorean Theorem

4. A *Pythagorean triple* is a triple (a, b, c) of positive integers such that $a^2 + b^2 = c^2$. A Pythagorean triple (a, b, c) is *primitive* if a, b, and c have no common factor. The tablet *Plimpton 322* indicates that the ancient Babylonians discovered the following method for generating all primitive Pythagorean triples. Start with relatively prime (i.e., no common factors) positive integers u and v, $u > v$, and then define $a = u^2 - v^2$, $b = 2uv$, and $c = u^2 + v^2$.

(a) Verify that (a, b, c) is a Pythagorean triple.

(b) Verify that a, b, and c are all even if u and v are both odd.

(c) Verify that (a, b, c) is a primitive Pythagorean triple in case one of u and v is even and the other is odd.

Every Pythagorean triple (a, b, c) with b even is generated by this Babylonian process. The proof of that fact is significantly more difficult than the exercises above but can be found in most modern number theory books.

5. The ancient Egyptians had a well-known interest in pyramids. According to the *Moscow papyrus*, they developed the following formula for the volume of a truncated pyramid with square base:

$$V = \frac{h}{3}(a^2 + ab + b^2).$$

In this formula, the base of the pyramid is an $a \times a$ square, the top is a $b \times b$ square and the height of the truncated pyramid (measured perpendicular to the base) is h. One fact you learned in high school geometry is that that volume of a pyramid is one-third the area of the base times the height. Use that fact along with some high school geometry and algebra to verify that the Egyptian formula is exactly correct.

6. Explain how to complete the following constructions using only compass and straightedge. (You probably learned to do this in high school.)
- **(a)** Given a line segment $\overline{AB}$, construct the perpendicular bisector of $\overline{AB}$.
- **(b)** Given a line ℓ and a point P not on ℓ, construct a line through P that is perpendicular to ℓ.
- **(c)** Given an angle $\angle BAC$, construct the angle bisector.

7. Can you prove the following assertions using only Euclid's postulates and common notions? Explain your answer.
- **(a)** Every line has at least two points lying on it.
- **(b)** For every line there is at least one point that does not lie on the line.
- **(c)** For every pair of points $A \neq B$, there is only one line that passes through A and B.

8. Find the first of Euclid's proofs in which he makes use of his Fifth Postulate.

9. A *rhombus* is a quadrilateral in which all four sides have equal lengths. The *diagonals* are the line segments joining opposite corners. Use the first five Propositions of Book I of the *Elements* to show that the diagonals of a rhombus divide the rhombus into four congruent triangles.

10. A *rectangle* is a quadrilateral in which all four angles have equal measures. (Hence they are all right angles.) Use the propositions in Book I of the *Elements* to show that the diagonals of a rectangle are congruent and bisect each other.

11. The following well-known argument illustrates the danger in relying too heavily on diagrams.[3] Find the flaw in the "Proof." (The proof uses familiar high school notation that will be explained later in this textbook. For example, $\overline{AB}$ denotes the segment from A to B and $\overleftrightarrow{AB}$ denotes the line through points A and B.)

False Proposition. *If $\triangle ABC$ is any triangle, then side $\overline{AB}$ is congruent to side $\overline{AC}$.*

Proof. Let ℓ be the line that bisects the angle $\angle BAC$ and let G be the point at which ℓ intersects $\overline{BC}$. Either ℓ is perpendicular to $\overline{BC}$ or it is not. We give a different argument for each case.

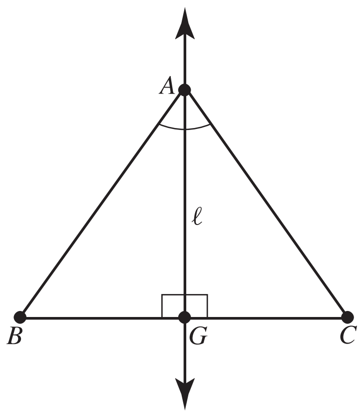

FIGURE 1.7: One possibility: the angle bisector is perpendicular to the base

Assume, first, that ℓ is perpendicular to $\overline{BC}$ (Figure 1.7). Then $\triangle AGB \cong \triangle AGC$ by Angle-Side-Angle and therefore $\overline{AB} \cong \overline{AC}$.

Now suppose ℓ is not perpendicular to $\overline{BC}$. Let m be the perpendicular bisector of $\overline{BC}$ and let M be the midpoint of $\overline{BC}$. Then m is perpendicular to $\overleftrightarrow{BC}$ and ℓ is not, so m is not equal to ℓ and m is not parallel to ℓ. Thus ℓ and m must intersect at a point D. Drop perpendiculars from D to the lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{AC}$ and call the feet of those perpendiculars E and F, respectively.

There are three possibilities for the location of D: either D is inside $\triangle ABC$, D is on $\triangle ABC$, or D is outside $\triangle ABC$. The three possibilities are illustrated in Figure 1.8.

Consider first the case in which D is on $\triangle ABC$. Then $\triangle ADE \cong \triangle ADF$ by Angle-Angle-Side, so $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Also $\overline{BD} \cong \overline{CD}$ since

[3]This fallacy is apparently due to W. W. Rouse Ball (1850–1925) and first appeared in the original 1892 edition of [3].

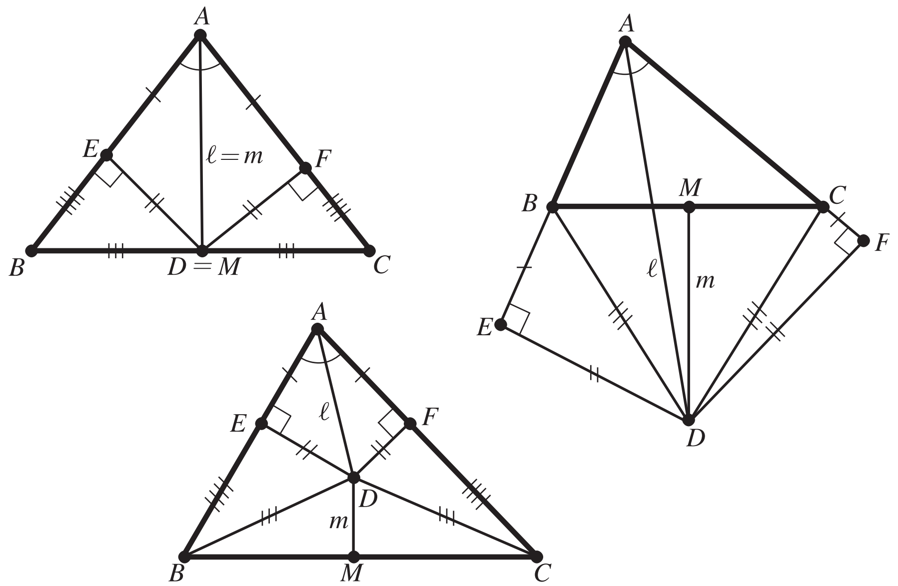

FIGURE 1.8: Three possible locations for D in case the angle bisector is not perpendicular to the base

$D = M$ is the midpoint of $\overline{BC}$. It follows from the Hypotenuse-Leg Theorem[4] that $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$. Hence $\overline{AB} \cong \overline{AC}$ by addition.

Next consider the case in which D is inside $\triangle ABC$. We have $\triangle ADE \cong \triangle ADF$ just as before, so again $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Also $\triangle BMD \cong \triangle CMD$ by Side-Angle-Side and hence $\overline{BD} \cong \overline{CD}$. Applying the Hypotenuse-Leg Theorem again gives $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$ as before. It follows that $\overline{AB} \cong \overline{AC}$ by addition.

Finally consider the case in which D is outside $\triangle ABC$. Once again we have $\triangle ADE \cong \triangle ADF$ by Angle-Angle-Side, so again $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Just as before, $\triangle BMD \cong \triangle CMD$ by Side-Angle-Side and hence $\overline{BD} \cong \overline{CD}$. Applying the Hypotenuse-Leg Theorem gives $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$. It then follows that $\overline{AB} \cong \overline{AC}$, this time by subtraction. □

[4]The Hypotenuse-Leg Theorem states that if the hypotenuse and leg of one right triangle are congruent to the corresponding parts of a second right triangle, then the triangles are congruent. It is a correct theorem; this is not the error in the proof.

CHAPTER 2

Axiomatic Systems and Incidence Geometry

Over the years that have passed since he wrote his *Elements*, Euclid's program for organizing geometry has been refined into what is called an *axiomatic system*. The basic structure of a modern axiomatic system was inspired by Euclid's method of organization, but there are several significant ways in which an axiomatic system differs from Euclid's scheme.

This chapter examines the various parts of an axiomatic system and explains their relationships. Those relationships are illustrated by a fundamental example known as incidence geometry. By doing a preliminary study of axioms and relations among them in the simple, uncomplicated setting of incidence geometry we are better able to understand how an axiomatic system works and what it means to say that one axiom is independent of some others. This lays the groundwork for the presentation of plane geometry as an axiomatic system in Chapter 3 and also prepares the way for later chapters where it is proved that Euclid's fifth postulate is independent of his other postulates.

An important feature of the axiomatic method is proof. The chapter contains a review of some basic principles used in the construction of proofs and also sets out the distinctive style of written proof that will used in this book. Incidence geometry provides a convenient setting in which to practice some of the proof-writing skills that will be required later.

2.1 THE STRUCTURE OF AN AXIOMATIC SYSTEM

The parts of an axiomatic system are undefined terms, definitions, axioms, theorems, and proofs. We will examine each of these parts separately.

Undefined and defined terms

The first part of an axiomatic system is a list of *undefined terms*. These are the technical words that will be used in the subject. Euclid attempted to define all his terms, but we now recognize that it is not possible to achieve the goal of defining all the terms. A standard dictionary appears to contain a definition of every word in a language, but there will inevitably be some circularity in the definitions. Rather than attempting to define every term we will use, we simply take certain key words to be undefined and work from there.

In geometry, we usually take such words as *point* and *line* to be undefined. In other parts of mathematics, the words *set* and *element of* are often undefined. When the real numbers are treated axiomatically, the term *real number* itself is sometimes undefined.

Even though some words are left undefined, there is still a place for definitions and defined terms in an axiomatic system. The aim is to start with a minimal number of undefined terms and then to define other technical words using the original undefined terms and previously defined terms. One role of definitions is just to allow statements to be made concisely. For example, we will define three points to be collinear if there is one line such that all three points lie on that line. It is much more clear and concise to say that three points are noncollinear than it is to say that there does not exist a single line such that all three points lie on that line. Another function of definitions is to identify and highlight key structures and concepts.

Axioms

The second part of an axiomatic system is a list of *axioms*. The words *axiom* and *postulate* are used interchangeably and mean exactly the same thing in this book.[1] The axioms are statements that are accepted without proof. They are where the subject begins. Everything else in the system should be logically deduced from them.

All relevant assumptions are to be stated in the axioms and the only properties of the undefined terms that may be used in the subsequent development of the subject are those that are explicitly spelled out in the axioms. Hence we will allow ourselves to use those and only those properties of points and lines that have been stated in our axioms—any other properties or facts about points and lines that we know from our intuition or previous experience are not to be used until and unless they have been proven to follow from the axioms.

One of the goals of this course is to present plane geometry as an axiomatic system. This will require a much more extensive list of axioms than Euclid used. The reason for this is that we must include *all* the assumptions that will be needed in the proofs and not allow ourselves to rely on diagrams or any intuitive but unstated properties of points and lines the way Euclid did.

Theorems and proofs

The final part, usually by far the largest part, of an axiomatic system consists of the theorems and their proofs. Again there are two different words that are used synonymously: the words *theorem* and *proposition* will mean the same thing in this course.[2] In this third part of an axiomatic system we work out the logical consequences of the axioms.

Just as in Euclid's *Elements*, there is a strict logical organization that applies. The first theorem is proved using only the axioms. The second theorem is proved using the first theorem together with the axioms, and so on.

Later in the chapter we will have much more to say about theorems and proofs as well as the rules of logic that are to be used in proofs.

---

[1] Generally *postulate* will be used when a particular assumption is being stated or referred to by name, while the word *axiom* will be used in a more generic sense to refer to unproven assumptions.

[2] There are also two other words that are used for theorem. A *lemma* is a theorem that is stated as a step toward some more important result. Usually a lemma is not an end in itself but is used as a way to organize a complicated proof by breaking it down into steps of manageable size. A *corollary* is a theorem that can be quickly and easily deduced from a previously stated theorem.

Interpretations and models

In an axiomatic system the undefined terms do not in themselves have any definite meaning, except what is explicitly stated in the axioms. The terms may be interpreted in any way that is consistent with the axioms. An *interpretation* of an axiomatic system is a particular way of giving meaning to the undefined terms in that system. An interpretation is called a *model* for the axiomatic system if the axioms are correct (true) statements in that interpretation. Since the theorems in the system were all logically deduced from the axioms and nothing else, we know that all the theorems will automatically be correct and true statements in any model.

We say that a statement in our axiomatic system is *independent* of the axioms if it is impossible to either prove or disprove the statement as a logical consequence of the axioms. A good way to show that a statement is independent of the axioms is to exhibit one model for the system in which the statement is true and another model in which it is false. As we shall see, that is exactly the way in which it was eventually shown that Euclid's fifth postulate is independent of Euclid's other postulates.

The axioms in an axiomatic system are said to be *consistent* if no logical contradiction can be derived from them. This is obviously a property we would want our axioms to have. Again it is a property that can be verified using models. If there exists a model for an axiomatic system, then the system must be consistent. The existence of a model for Euclidean geometry and thus the consistency of Euclid's postulates was taken for granted until the nineteenth century. Our study of geometry will repeat the historical pattern: We will first study various geometries as axiomatic systems and only address the questions of consistency and existence of models in Chapter 11.

2.2 AN EXAMPLE: INCIDENCE GEOMETRY

In order to clarify what an axiomatic system is, we study the important example of *incidence geometry*. For now we simply look at the axioms and various models for this system; a more extensive discussion of theorems and proofs in incidence geometry is delayed until later in the chapter.

Let us take the three words *point, line*, and *lie on* (as in "point P lies on line ℓ") to be our undefined terms. The word *incident* is also used in place of *lie on*, so the two statements "P lies on ℓ" and "P is incident with ℓ" mean the same thing. For that reason the axioms for this relationship are called *incidence axioms*. One advantage of the word incident is that it can be used symmetrically: We can say that P is incident with ℓ or that ℓ is incident with P; both statements mean exactly the same thing.

There are three incidence axioms. When we say (in the axiom statements) that *P and Q are distinct points*, we simply mean that they are not the same point.

Incidence Axiom 1. *For every pair of distinct points P and Q there exists exactly one line ℓ such that both P and Q lie on ℓ.*

Incidence Axiom 2. *For every line ℓ there exist at least two distinct points P and Q such that both P and Q lie on ℓ.*

Incidence Axiom 3. *There exist three points that do not all lie on any one line.*

The axiomatic system with the three undefined terms and the three axioms listed above is called *incidence geometry*. We usually also call a model for the axiomatic system *an incidence geometry* and an interpretation of the undefined terms is called *a geometry*. Before giving examples of incidence geometries, it is convenient to introduce a defined term.

Definition 2.2.1. Three points A, B, and C are *collinear* if there exists one line ℓ such that all three of the points A, B, and C all lie on ℓ. The points are *noncollinear* if there is no such line ℓ.

Using this definition we can give a more succinct statement of Incidence Axiom 3: *There exist three noncollinear points.*

EXAMPLE 2.2.2 Three-point geometry

Interpret *point* to mean one of the three symbols A, B, C; interpret *line* to mean a set of two points; and interpret *lie on* to mean "is an element of." In this interpretation there are three lines, namely $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$. Since any pair of distinct points determines exactly one line and no one line contains all three points, this is a model for incidence geometry. ■

Be sure to notice that this "geometry" contains just three points. It is an example of a *finite geometry*, which is a geometry that contains only a finite number of points. It is customary to picture such geometries by drawing a diagram in which the points are represented by dots and the lines by segments joining them. So the diagram for the three-point plane looks like a triangle (see Figure 2.1). Don't be misled by the diagram: the "points" on the sides of the triangle are *not* points in the three-point plane. The diagram is strictly schematic, meant to illustrate relationships, and is not to be taken as a literal picture of the geometry.

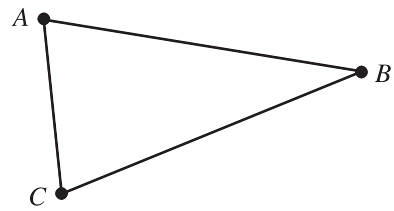

FIGURE 2.1: Three-point geometry

EXAMPLE 2.2.3 The three-point line

Interpret *point* to mean one of the three symbols A, B, C, but this time interpret *line* to mean the set of all points. This geometry contains only one line, namely $\{A, B, C\}$. In this interpretation Incidence Axioms 1 and 2 are satisfied, but Incidence Axiom 3 is not satisfied. Hence the three-point line is not a model for incidence geometry (Figure 2.2). ■

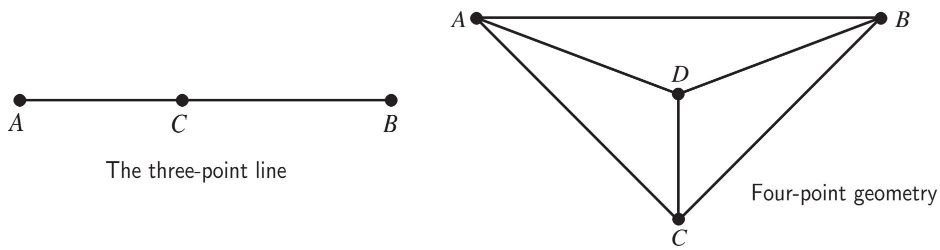

FIGURE 2.2: Two interpretations of the terms of incidence geometry

EXAMPLE 2.2.4 Four-point geometry

Interpret *point* to mean one of the four symbols A, B, C, D; interpret *line* to mean a set of two points and interpret *lie on* to mean "is an element of." In this interpretation there are six lines, namely $\{A, B\}$, $\{A, C\}$, $\{A, D\}$, $\{B, C\}$, $\{B, D\}$, and $\{C, D\}$. Since any pair of distinct points determines exactly one line and no one line contains three distinct points, this is a model for incidence geometry (Figure 2.2). ■

EXAMPLE 2.2.5 Five-point geometry

Interpret *point* to mean one of the five symbols A, B, C, D, E; interpret *line* to mean a set of two points and interpret *lie on* to mean "is an element of." In this interpretation there are ten lines. Again any pair of distinct points determines exactly one line and no one line contains three distinct points, so this is also a model for incidence geometry. ■

We could continue to produce n-point geometries for increasingly large values of n, but we will stop with these three because they illustrate all the possibilities regarding parallelism that will be studied in the next section.

EXAMPLE 2.2.6 The interurban

In this interpretation there are three *points*, namely the cities of Grand Rapids, Holland, and Muskegon (three cities in western Michigan). A *line* consists of a railroad line from one city to another. There is one railroad line joining each pair of distinct cities, for a total of three lines. Again, this is a model for incidence geometry. ■

EXAMPLE 2.2.7 Fano's geometry

Interpret *point* to mean one of the seven symbols A, B, C, D, E, F, G; interpret *line* to mean one of the seven three-point sets listed below and interpret *lie on* to mean "is an element of." The seven lines are

$$\{A, B, C\}, \{C, D, E\}, \{E, F, A\}, \{A, G, D\}, \{C, G, F\}, \{E, G, B\}, \{B, D, F\}.$$

All three incidence axioms hold in this interpretation, so Fano's geometry[3] is another model for incidence geometry (Figure 2.3). ■

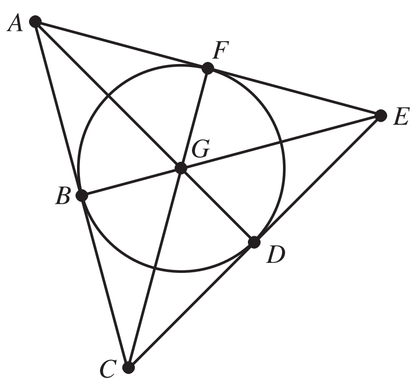

FIGURE 2.3: Fano's geometry

The illustration of Fano's Geometry (Figure 2.3) shows one of the lines as curved while the others are straight. It should be recognized that this is an artifact of the schematic diagram we use to picture the geometry and is not a difference in the lines themselves. A line is simply a set of three points; the curves in the diagram are meant to show visually which points lie together on a line and are not meant to indicate anything about straightness.

The examples of interpretations given so far illustrate the fact that the undefined terms in a given axiomatic system can be interpreted in widely different ways. No one of the models is preferred over any of the others. Notice that three-point geometry and the interurban are essentially the same; the names for the points and lines are different, but all the important relationships are the same. We could easily construct a correspondence from the set of

[3]Named for Gino Fano, 1871–1952.

points and lines of one model to the set of points and lines of the other model. The correspondence would preserve all the relationships that are important in the geometry (such as incidence). Models that are related in this way are called *isomorphic* models and a function between them that preserves all the geometric relationships is an *isomorphism*.

All the models described so far have been finite geometries. Of course the geometries with which we are most familiar are not finite. We next describe three infinite geometries.

EXAMPLE 2.2.8 The Cartesian plane

In this model a *point* is any ordered pair (x, y) of real numbers. A *line* is the collection of points whose coordinates satisfy a linear equation of the form $ax + by + c = 0$, where a, b, and c are real numbers and a and b are not both 0. More specifically, three real numbers a, b, and c, with a and b not both 0, determine the line ℓ consisting of all pairs (x, y) such that $ax + by + c = 0$. A point (x, y) is said to *lie on* the line if the coordinates of the point satisfy the equation. This is just the coordinate (or Cartesian) plane model for high school Euclidean geometry. It is also a model for incidence geometry. We will use the symbol $\mathbb{R}^2$ to denote the set of points in this model (Figure 2.4). ■

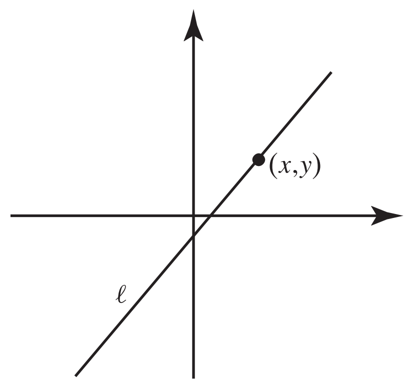

FIGURE 2.4: The Cartesian plane

EXAMPLE 2.2.9 The sphere

Interpret *point* to mean a point on the surface of a round 2-sphere in three-dimensional space. Specifically, a point is an ordered triple (x, y, z) of real numbers such that $x^2 + y^2 + z^2 = 1$. A *line* is interpreted to mean a great circle on the sphere and *lie on* is again interpreted to mean "is an element of." We will use the symbol $\mathbb{S}^2$ to denote the set of points in this model.

A *great circle* is a circle on the sphere whose radius is equal to that of the sphere. A great circle is the intersection of a plane through the origin in 3-space with the sphere. Two points on the sphere are *antipodal* (or opposite) if they are the two points at which a line through the origin intersects the sphere. Two given antipodal points on the sphere lie on an infinite number of different great circles; hence this geometry does not satisfy Incidence Axiom 1. If A and B are two points on the sphere that are not antipodal, then A and B determine a unique plane through the origin in 3-space and thus lie on a unique great circle. Hence "most" pairs of points determine a unique line in this geometry (Figure 2.5).

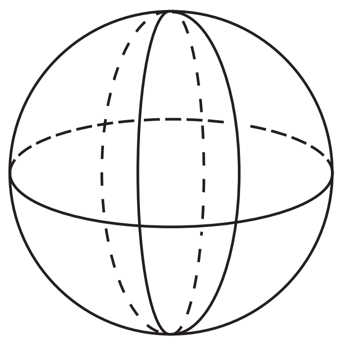
FIGURE 2.5: The sphere

Since this interpretation does not satisfy Incidence Axiom 1, it is not a model for incidence geometry. Note that Incidence Axioms 2 and 3 are correct statements in this interpretation. Another important observation about the sphere is that there are no parallel lines: any two distinct great circles on the sphere intersect in a pair of points. ■

EXAMPLE 2.2.10 The Klein disk

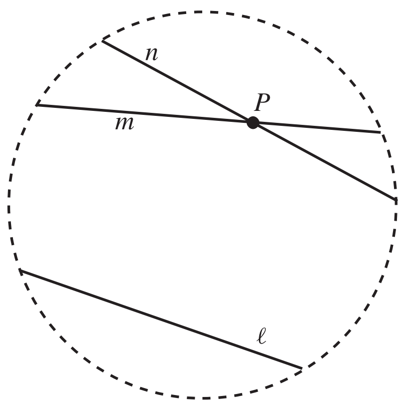

FIGURE 2.6: The Klein disk

Interpret *point* to mean a point in the Cartesian plane that lies inside the unit circle. In other words, a point is an ordered pair (x, y) of real numbers such that $x^2 + y^2 < 1$. A *line* is the part of a Euclidean line that lies inside the circle and *lie on* has its usual Euclidean meaning. This is a model for incidence geometry (Figure 2.6).

The Klein disk is an infinite model for incidence geometry, just like the familiar Cartesian plane is. (In this context *infinite* means that the number of points is unlimited, not that distances are unbounded.) The two models are obviously different in superficial ways. But they are also quite different with respect to some of the deeper relationships that are important in geometry. We illustrate this in the next section by studying parallel lines in each of the various geometries we have described. ■

2.3 THE PARALLEL POSTULATES IN INCIDENCE GEOMETRY

Next we investigate parallelism in incidence geometry. The purpose of the investigation is to clarify what it means to say that the Euclidean Parallel Postulate is independent of the other axioms of geometry.

We begin with a definition of the word parallel, which becomes our second defined term in incidence geometry. In high school geometry parallel lines can be characterized in many different ways, so you may recall several definitions of parallel. In the context of everything that is assumed in high school geometry those definitions are logically equivalent and can be used interchangeably. But we have made no assumptions, other than those stated in the incidence axioms, so we must choose one of the definitions and make it the official definition of parallel. We choose the simplest characterization: the lines do not intersect. That definition fits best because it can be formulated using only the undefined terms of incidence geometry. It is also Euclid's definition of what it means for two lines in a plane to be parallel (Definition 23, Appendix A). Obviously this would not be the right definition to use for lines in 3-dimensional space, but the geometry studied in this book is restricted to the geometry of the two-dimensional plane. Note that, according to this definition of parallel, a line is not parallel to itself.

Definition 2.3.1. Two lines ℓ and m are said to be *parallel* if there is no point P such that P lies on both ℓ and m. The notation for parallelism is $\ell \parallel m$.

There are three different parallel postulates that will be useful in this course. The first is called the Euclidean Parallel Postulate, even though it is not actually one of Euclid's postulates. We will see later that (in the right context) it is logically equivalent to Euclid's fifth postulate. This formulation of the Euclidean Parallel Postulate is often called *Playfair's Postulate* (see page 93).

Euclidean Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there is exactly one line m such that P lies on m and $m \parallel \ell$.*

There are other possibilities besides the Euclidean one. We state two of them.

Elliptic Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there is no line m such that P lies on m and $m \parallel \ell$.*

Hyperbolic Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there are at least two lines m and n such that P lies on both m and n and both m and n are parallel to ℓ.*

These are not new axioms for incidence geometry. Rather they are additional statements that may or may not be satisfied by a particular model for incidence geometry. We illustrate with several examples.

EXAMPLE 2.3.2 Parallelism in three-, four-, and five-point geometries

In three-point geometry any two lines intersect. Therefore there are no parallel lines and this model satisfies the Elliptic Parallel Postulate.

Each line in four-point geometry is disjoint from exactly one other line. Thus, for example, the line $\{A, B\}$ is parallel to the line $\{C, D\}$ and no others. There is exactly one parallel line that is incident with each point that does not lie on $\{A, B\}$. Since the analogous statement is true for every line, four-point geometry satisfies the Euclidean Parallel Postulate.

Consider the line $\{A, B\}$ in five point geometry and the point C that does not lie on $\{A, B\}$. Observe that C lies on two different lines, namely $\{C, D\}$ and $\{C, E\}$, that are both parallel to $\{A, B\}$. Since this happens for every line and for every point that does not lie on that line, five-point geometry satisfies the Hyperbolic Parallel Postulate. ■

EXAMPLE 2.3.3 Parallelism in the Cartesian plane, sphere, and Klein disk

The Euclidean Parallel Postulate is true in the Cartesian plane $\mathbb{R}^2$. The fact that the Cartesian plane satisfies the Euclidean Parallel Postulate is probably familiar to you from your high school geometry course.

The sphere $\mathbb{S}^2$ satisfies the Elliptic Parallel Postulate. The reason for this is simply that there are no parallel lines on the sphere (any two great circles intersect).

The Klein disk satisfies the Hyperbolic Parallel Postulate. The fact that the Klein disk satisfies the Hyperbolic Parallel Postulate is illustrated in Figure 2.6. ■

Conclusion

We can conclude from the preceding examples that each of the parallel postulates is independent of the axioms of incidence geometry. For example, the fact that there are some models for incidence geometry that satisfy the Euclidean Parallel Postulate and there are other models that do not shows that neither the Euclidean Parallel Postulate nor its negation can be proved as a theorem in incidence geometry. The examples make it clear that it would be fruitless to try to prove any of the three parallel postulates in incidence geometry.

This is exactly how it was eventually shown that Euclid's fifth postulate is independent of his other postulates. Understanding that proof is one of the major goals of this course. Later in the book we will construct two models for geometry, both of which satisfy all of Euclid's assumptions other than his fifth postulate. One of the models satisfies Euclid's fifth postulate while the other does not. (It satisfies the Hyperbolic Parallel Postulate.) This shows that it is impossible to prove Euclid's fifth postulate using only Euclid's other postulates and assumptions. It will take us most of the course to fully develop those models. One of the reasons it is such a difficult task is that we have to dig out *all* of Euclid's assumptions—not just the assumptions stated in the postulates, but all the unstated ones as well.

2.4 AXIOMATIC SYSTEMS AND THE REAL WORLD

An axiomatic system, as defined in this chapter, is obviously just a refinement of Euclid's system for organizing geometry. It should be recognized, however, that these refinements have profound implications for our understanding of the place of mathematics in the world.

The ancient Greeks revolutionized geometry by making it into an abstract discipline. Before that time, mathematics and geometry had been closely tied to the physical world. Geometry was the study of one aspect of the real world, just like physics or astronomy. In fact, the word *geometry* literally means "to measure the earth."

Later Greek geometry, on the other hand, is about relationships between ideal, abstract objects. In this view, geometry is not just about the physical world in which we live our everyday lives, but it also gives us information about an ideal world of pure forms. In the view of Greek philosophers such as Plato, this ideal world was, if anything, more real than the physical world of our existence. The relationships in the ideal world are eternal and pure. The Greeks presumably thought of a postulate as a statement about relationships that really pertain in that ideal world. The postulates are true statements that can be accepted without proof because they are self-evident truths about the way things really are in the ideal world. So the Greeks distanced geometry from the physical world by making it abstract, but at the same time they kept it firmly rooted in the real world where it could give them true and reliable information about actual spatial relationships.

We have no direct knowledge of how Euclid himself understood the significance of his geometry. All we know about his thinking is what we find written in the *Elements* and those books are remarkably terse by modern standards. But Euclid lived approximately 100 years after Plato, so it seems reasonable to assume that he was influenced by Plato's ideas. In any event, it is quite clear from reading the *Elements* that Euclid thought of geometry as being about real things and that is precisely why he felt free to use intuitively obvious facts about points and lines in his proofs even though he had not stated these facts as axioms.

The view of geometry as an axiomatic system (as described in this chapter) moves us well beyond the Platonic view. In our effort to spell out completely what our assumptions are, we have been led to make geometry much more relative and detached from reality. We do not apply the terms true or false to the axioms in any absolute sense. An axiom is simply a statement that may be true or may be false in any particular situation, it just depends on how we choose to interpret the undefined terms. Thus our efforts to introduce abstraction and rigor into geometry have led us to drain the meaning out of such everyday terms as point and line. Since the words can now mean just about anything we want them to, we must wonder whether they any longer have any real content.

The naïve view is that geometry is the study of space and spatial relationships. We usually think of geometry as a science that gives us true and reliable information about the world in which we live. The view of geometry as an axiomatic system detached from the real world is a bit disturbing to most of us.

Some mathematicians have promoted the view that mathematics is just a logical game in which we choose an arbitrary set of axioms and then see what we can deduce using the rules of logic. Most professional mathematicians, however, have a profound sense that the mathematics they study is about real things. The fact that mathematics has such incredibly powerful and practical applications is evidence that it is much more than a game.

It is surprisingly difficult to resolve the kinds of philosophical issues that are raised by these observations. The mathematical community's thinking on these matters has evolved over time and there have been several amazing revolutions in the conventional understanding of what the correct views should be. Those views will be explored in this book as the historical development of geometry unfolds. We do not attempt to give definitive answers; instead we simply raise the questions and encourage the reader to think about them. In particular, the following questions should be recognized and should be kept in mind as the development of geometry is worked out in this book.

1. Are the theorems of geometry true statements about the world in which we live?
2. What physical interpretation should we attach to the terms *point* and *line*?
3. What are the axioms that describe the geometry of the space in which we live?
4. Is it worthwhile to study arbitrary axiom systems or should we restrict our attention to just those axiom systems that appear to describe the real world?

At this point you might be asking yourself why it would be thought desirable to make mathematics so abstract and therefore to get into the kind of difficult issues that have been raised here. That is one question we can answer. The answer is that abstraction is precisely what gives mathematics its power. By identifying certain key features in a given situation, listing exactly what it is about those features that is to be studied, and then studying them in an abstract setting detached from the original context, we are able to see that the same kinds of relationships hold in many apparently different contexts. We are able to study the important relationships in the abstract without a lot of other irrelevant information cluttering up the picture and obscuring the underlying structure. Once things have been clarified in this way, the kind of logical reasoning that characterizes mathematics becomes an incredibly powerful and effective tool. The history of mathematics is full of examples of surprising practical applications of mathematical ideas that were originally discovered and developed by people who were completely unaware of the eventual applications.

EXERCISES 2.4

1. It is said that Hilbert once illustrated his contention that the undefined terms in a geometry should not have any inherent meaning by claiming that it should be possible to replace *point* by *beer mug* and *line* by *table* in the statements of the axioms. Consider three friends sitting around one table. Each person has one beer mug. At the moment all the beer mugs are resting on the table. Suppose we interpret point to mean beer mug, line to mean the table, and lie on to mean resting on. Is this a model for incidence geometry? Explain. Is this interpretation isomorphic to any of the examples in the text?
2. One-point geometry contains just one point and no line. Which incidence axioms does one-point geometry satisfy? Explain. Which parallel postulates does one-point geometry satisfy? Explain.
3. Consider a small mathematics department consisting of Professors Alexander, Bailey, Curtis, and Dudley with three committees: curriculum committee, personnel committee, and social committee. Interpret *point* to mean a member of the department, interpret *line* to be a departmental committee, and interpret *lie on* to mean that the faculty member is a member of the specified committee.
   (a) Suppose the committee memberships are as follows: Alexander, Bailey, and Curtis are on the curriculum committee; Alexander and Dudley are on the personnel

committee; and Bailey and Curtis are on the social committee. Is this a model for Incidence Geometry? Explain.

(b) Suppose the committee memberships are as follows: Alexander, Bailey and Curtis are on the curriculum committee; Alexander and Dudley are on the personnel committee; and Bailey and Dudley are on the social committee. Is this a model for incidence geometry? Explain.

(c) Suppose the committee memberships are as follows: Alexander and Bailey are on the curriculum committee, Alexander and Curtis are on the personnel committee, and Dudley and Curtis are on the social committee. Is this a model for incidence geometry? Explain.

4. A three-point geometry is an incidence geometry that satisfies the following additional axiom: *There exist exactly three points.*

(a) Find a model for three-point geometry.

(b) How many lines does any model for three-point geometry contain? Explain.

(c) Explain why any two models for three-point geometry must be isomorphic. (An axiomatic system with this property is said to be *categorical*.)

5. Interpret *point* to mean one of the four vertices of a square, *line* to mean one of the sides of the square, and *lie on* to mean that the vertex is an endpoint of the side. Which incidence axioms hold in this interpretation? Which parallel postulates hold in this interpretation?

6. Draw a schematic diagram of five-point geometry (see Example 2.2.5).

7. Which parallel postulate does Fano's geometry satisfy? Explain.

8. Which parallel postulate does the three-point line satisfy? Explain.

9. Under what conditions could a geometry satisfy more than one of the parallel postulates? Explain. Could an incidence geometry satisfy more than one of the parallel postulates? Explain.

10. Consider a finite model for incidence geometry that satisfies the following additional axiom: *Every line has exactly three points lying on it.* What is the minimum number of points in such a geometry? Explain your reasoning.

11. Find a finite model for Incidence Geometry in which there is one line that has exactly three points lying on it and there are other lines that have exactly two points lying on them.

12. Find interpretations for the words *point*, *line*, and *lie on* that satisfy the following conditions.

(a) Incidence Axioms 1 and 2 hold, but Incidence Axiom 3 does not.

(b) Incidence Axioms 2 and 3 hold, but Incidence Axiom 1 does not.

(c) Incidence Axioms 1 and 3 hold, but Incidence Axiom 2 does not.

13. For any interpretation of incidence geometry there is a *dual* interpretation. For each point in the original interpretation there is a line in the dual and for each line in the original there is point in the dual. A point and line in the dual are considered to be incident if the corresponding line and point are incident in the original interpretation.

(a) What is the dual of the three-point plane? Is it a model for incidence geometry?

(b) What is the dual of the three-point line? Is it a model for incidence geometry?

(c) What is the dual of four-point geometry? Is it a model for incidence geometry?

(d) What is the dual of Fano's geometry?

2.5 THEOREMS, PROOFS, AND LOGIC

We now take a more careful look at the third part of an axiomatic system: the theorems and proofs. Both theorems and proofs require extra care. Most of us have enough experience with mathematics to know that the ability to write good proofs is a skill that must be learned, but we often overlook the fact that a necessary prerequisite to good proof writing is good statements of theorems.

A major goal of this course is to teach the art of writing proofs and it is not expected that the reader is already proficient at it. The main way in which one learns to write proofs is by actually writing them, so the remainder of the book will provide lots of opportunities for practice. This section simply lays out a few basic principles and then those principles will be put to work in the rest of the course. The brief introduction provided in this section will not make you an instant expert at writing proofs, but it will equip you with the basic tools you need to get started. You should refer back to this section as necessary in the remainder of the course.

Mathematical language

An essential step on the way to the proof of a theorem is a careful statement of the theorem in clear, precise, and unambiguous language. To illustrate this point, consider the following proposition in incidence geometry.

Proposition 2.5.1. *Lines that are not parallel intersect in one point.*

Compare that statement with the following.

Restatement. *If ℓ and m are two distinct lines that are not parallel, then there exists exactly one point P such that P lies on both ℓ and m.*

Both are correct statements of the same theorem. But the second statement is much better, at least as far as we are concerned, because it clearly states where the proof should begin (with two lines ℓ and m such that $\ell \neq m$ and $\ell \nparallel m$) and where it should end (with a point P that lies on both ℓ and m). This provides the framework within which we can build a proof. Contrast that with the first statement. In the first statement it is much less clear how to begin a proof. In fact it is not possible to construct a proof until we have at least mentally translated the first statement into language that is closer to that in the second statement. When we start to do that we realize that the first statement is not precise enough. For example, it does not clearly say whether it is an assertion about two (or more?) particular lines or whether it is making an assertion that applies to all lines. Writing good proofs requires clear thinking and clear thinking begins with careful statements.

Statements

In mathematics, the word *statement* refers to any assertion that can be classified as either true or false (but not both). Here is an example: Dan is tiny. The statements of geometry often involve assertions that objects (such as points or lines) satisfy certain conditions (such as parallelism). Such statements must be preceded by definitions of the terms used. For example, it is not possible to determine if the assertion *Dan is tiny* is true or false unless we have a precise definition of what *tiny* means in this context. It is obvious that the word *tiny* might mean one thing in one situation (for example, in microbiology) and something completely different in another context (such as astronomy). So we would need a definition of the form, "A person x is said to be *tiny* if ...." Once you have that definition, you can check to see whether or not a particular person named Dan satisfies the conditions in the definition.

Simple statements can be combined into compound statements using the words *and* and *or*. The use of *and* is easy to understand; it means that both the statements are true. The use of *or* in mathematics differs somewhat from the way the word is used in ordinary language. In mathematics *or* is always used in a nonexclusive way; it means that one or the other of the statements is true and allows the possibility that both are true. Consider the statement *Joan is old or Joan is rich*. As used in ordinary everyday English, this statement allows the possibility that Joan is both old and rich. Contrast that with the following statement: *Either you are for me or you are against me*. We understand from the tone of the statement that it means you are one or the other but not both. Thus the word *or* is ambiguous in ordinary English and its exact meaning depends on the context. Mathematical language eliminates such ambiguities and the word *or* always has the nonexclusive meaning when it is used in a mathematical statement.

We will often want to negate statements. Specifically, given one statement we will want to write down a second statement that asserts the opposite of the first. There is a sense in which it is easy to negate a statement: simply say, "It is not the case that ...." But this is not helpful. It is much more useful to observe that negation interchanges *and* and *or*. In other words, we have the following laws (for any statements S and T).

not (S and T) = (not S) or (not T)

not (S or T) = (not S) and (not T)

For example, if it is not true that $x > 0$ and $x < 1$, then either $x \leqslant 0$ or $x \geqslant 1$. The two rules stated above are known as *DeMorgan's laws*.

Propositional functions

An assertion such as "$x > 0$" does not, by itself, qualify as a statement in the technical sense defined above because it is neither true nor false until a value has been assigned to x. The assertion is really a function whose domain consists of numbers x and whose range consists of the values "True" and "False." Such a function is called a *propositional function*.[4] A specific example of a propositional function is the function defined by the formula $P(x) = (x > 0)$. The domain of P is the set of real numbers and the range is the set $\{\text{True, False}\}$. Thus, for example, $P(0) = \text{False}$ and $P(1) = \text{True}$. Propositional functions can have more than one independent variable, as in $Q(x, y) = (x > y)$. In this example, $Q(1, 1) = \text{False}$ and $Q(1, 0) = \text{True}$. This is precisely the kind of assertion we will encounter in the geometry course. For example, the assertion $\ell \parallel m$ should be interpreted as a propositional function whose domain is the set of ordered pairs of lines.

Conditional statements

A *conditional statement* is a compound statement of the form "If ..., then ..." in which the first set of dots represents a statement called the *hypothesis* (or *antecedent*) and the second set of dots represents a statement called the *conclusion* (or *consequent*). A *theorem* is a conditional statement that has been proved true. A conditional statement may be either true or false, but it is not called a theorem unless it is true and has been (or can be) proved. This means that there are no false theorems, just statements that have the form of theorems but turn out not to be theorems. Here is a good rule to follow: *Every theorem should be stated in if-then form.*[5]

Another way in which a conditional statement can be written is this: hypothesis implies conclusion. Notation: Hypothesis $\Rightarrow$ Conclusion.

A theorem does not assert that the conclusion is true without the hypothesis, only that the conclusion is true if the hypothesis is. If H and C are propositional functions, the

[4] In some contexts the word "proposition" is used for what we are calling a "statement." That is the origin of the terminology *propositional* function. In this book the word "proposition" is almost always used as a synonym for "theorem."

[5] As is the case with most rules, this one allows some exceptions. Here is a well-known theorem from calculus: *π is irrational*. In this case the hypotheses are hidden in the definition of π. While such a statement does qualify as a theorem, it is not a model we should adopt for this course. The practice of stating the hypotheses explicitly will serve us well as we learn to write proofs. The example does illustrate the fact that theorem statements are context dependent and there are often unstated hypotheses that are assumed in a given setting.

statement $H \Rightarrow C$ means that C is true in every case in which H is true. An example from high school algebra is the theorem, *If* $x < 1$, *then* $x < 2$. This is a true statement because the propositional function $x < 2$ is true for every x for which the propositional function $x < 1$ is true. The conditional statement $H \Rightarrow C$ really just rules out the possibility that H is true while C is false. Here is a more subtle example: *If x is a real number and* $x^2 < 0$, *then* $x = 3$. This is considered to be a logically correct statement because the conclusion is true of every x for which the hypothesis is true. (There are no x for which the hypothesis is true.) In a case like this we usually say that the theorem is *vacuously true* since the theorem is true only because there is no way the hypotheses can be true.

Negating a conditional statement requires clear thinking. The statement $P \Rightarrow Q$ means that Q is true whenever P is. The negation of $P \Rightarrow Q$ is the assertion that it is possible for P to be true while Q is false. Note that the negation of a conditional statement is not another conditional statement. Consider the following example of a proposed theorem: *If x is irrational, then x^2 is irrational.* This statement is false[6] because there are some irrational numbers whose square is rational [e.g., $\sqrt{2}$ is irrational while $(\sqrt{2})^2$ is rational]. It is true that there are some irrational numbers whose squares are irrational, but it takes only one example to show that the conditional statement is false. For this reason we normally demonstrate that a conditional statement is false by producing a counterexample.

Converse and contrapositive

For every conditional statement there are two related statements called the *converse* and the *contrapositive*. The converse of $P \Rightarrow Q$ is $Q \Rightarrow P$ and the contrapositive is *not* $Q \Rightarrow$ *not* P. The converse to a conditional statement is an entirely different statement; the fact that $P \Rightarrow Q$ is a theorem tells us nothing about whether or not the converse is a theorem. On the other hand, the contrapositive is logically equivalent to the original statement. Here is a simple example: If $x = 2$, then $x^2 = 4$. This is a correct theorem. Its converse, however, is not correct. ($x^2 = 4 \not\Rightarrow x = 2$.) The contrapositive is this: If $x^2 \neq 4$, then $x \neq 2$. The contrapositive is a correct statement and is just a negative way of restating the original theorem.

Consider another simple example: If $x = 0$, then $x^2 = 0$. This time both the statement and its converse are true. Such statements are called *biconditional* statements and the phrase "if and only if" (abbreviated iff) is used to indicate that the implication goes both ways. In other words, P iff Q (or $P \Leftrightarrow Q$) means $P \Rightarrow Q$ and $Q \Rightarrow P$. Thus we could say that $x = 0$ iff $x^2 = 0$. An if-and-only-if statement is really two theorems in one and the proof should reflect this; there should normally be separate proofs of each of the two implications.

Truth tables

There is a sense in which the fact that a conditional statement and its contrapositive are equivalent is obvious, but it can be confusing to explain the equivalence because negations are piled on negations. A simple device that can be used to explain the equivalence is a *truth table*. This is a good way to think of it because the truth table also sheds light on the meaning of the conditional statement itself.

Consider the statement $H \Rightarrow C$. The hypothesis and the conclusion can each be either true or false. Thus there are four possibilities for H and C and the statement $H \Rightarrow C$ is true in three of the four cases. The various possibilities are shown in the following truth table.

---

[6]The real theorem is this: *If x is rational, then x^2 is rational.*

H	C	$H \Rightarrow C$
True	True	True
True	False	False
False	True	True
False	False	True

It is the second half of the table that often confuses beginners; these are the cases in which the theorem is vacuously true. Since the conditional statement is true in three of the four cases, a proof is simply an argument that rules out the fourth possibility. If we now expand the table to include the negations of H and C as well as the contrapositive of the theorem we see that the contrapositive is true exactly when the theorem is. (The third and sixth columns are identical.)

H	C	$H \Rightarrow C$	*not* H	*not* C	*not* $C \Rightarrow$ *not* H
True	True	True	False	False	True
True	False	False	False	True	False
False	True	True	True	False	True
False	False	True	True	True	True

Sometimes it is more convenient to prove the contrapositive of a theorem than it is to prove the theorem itself. This is perfectly legitimate because the original statement is logically equivalent to the contrapositive.

Quantifiers

One of the distinctions that must be made clear is whether you are asserting that *every* object of a certain type satisfies a condition or whether you are simply asserting that there is one that does. This is specified through the use of *quantifiers*. There are two quantifiers: the existential quantifier (written $\exists$) and the universal quantifier (written $\forall$).

As the name implies, the existential quantifier asserts that something exists. Example: There exists a point P such that P does not lie on ℓ. If this statement occurs as part of the conclusion of a theorem then your proof need only exhibit one point P that does not lie on ℓ. It is important to be clear about this so that you don't find yourself trying to prove, for example, that every point does not lie on ℓ. The universal quantifier, on the other hand, is used to say that some property holds for all objects in a certain class. Example: For every point P not on ℓ, the distance from P to ℓ is positive. In this case the strategy of the proof would be completely different; instead of exhibiting one particular P for which the distance is positive we should give a universal proof that works for every P not on ℓ. Such a proof would begin with a statement such as this: "Let P be a point that does not lie on ℓ." The proof should then go on to demonstrate that the distance from P to ℓ is positive, using only the fact that P does not lie on ℓ.

The universal quantifier can sometimes be used in place of a conditional statement. For example, the following statements are logically equivalent: If $x > 2$, then $x^2 > x + 2$ and, For every $x > 2$, $x^2 > x + 2$.

Negation interchanges the two quantifiers. For example, consider this statement: Every angle is acute. Stated more precisely, it says, for every angle α, α is acute. The negation is, there exists an angle α such that α is not acute. Here is another example: There exists a point that does not lie on ℓ. The negation is this: Every point lies on ℓ. Better yet is this statement: For every point P, P lies on ℓ. The rules for negating quantified statements are new versions of DeMorgan's Laws:

not $(\forall x P(x)) = \exists x$ (not $P(x)$)
not $(\exists x P(x)) = \forall x$ (not $P(x)$)

Uniqueness

The word *unique* is often used in connection with the existential quantifier. For example, here is a statement that is important in geometry: For every line ℓ and for every point P there is a unique line m such that P lies on m and m is perpendicular to ℓ. The word *unique* in this statement indicates that there is exactly one line m satisfying the stated conditions. A proof of this assertion should have two parts. First, there should be a proof that there is a line m satisfying the conditions and, second, a proof that there cannot be two different lines m and n satisfying the conditions. The usual strategy for the second half of the proof is to start with the assumption that m and n are lines that satisfy the property and then to prove that m and n must, in fact, be equal to each other. The symbol ! is used to indicate uniqueness; so the notation "$\exists$! ..." should be read "there exists a unique ...."

Writing proofs

A proof consists of a sequence of steps that lead us logically from the hypothesis to the conclusion. Each step should be justified by a reason. There are six kinds of reasons that can be given:

- by hypothesis
- by axiom
- by previous theorem
- by definition
- by an earlier step in this proof
- by one of the rules of logic

In this course we plan to assume much of what you already know about the algebra of real numbers. Thus the first several kinds of reasons listed above will often be stated as "by properties of real numbers" or "by algebra."

At the beginning of the course we will follow Euclid's practice of writing the reasons in parentheses after the statements. We will eventually drop that style as we develop more proficiency at writing and reading proofs. But for now, it is important to spell out all your justifications.

In high school you may have learned to write your proofs in two columns, with the statements in one column and the reasons in another. We will not do that in this course, not even at the outset, because we are aiming to write proofs that can be read by fellow humans; in order to facilitate this, the proofs should be written in ordinary paragraph form. For the same reason we will not follow the high school custom of numbering the statements in a proof.

It is helpful to distinguish between the proof itself and the written argument that is used to communicate the proof to other people. The proof is a sequence of logical steps that lead from the hypothesis to the conclusion. The written argument lays out those steps for the reader, and the writer has an obligation to write them in a way that the reader can understand without undue effort. So the written proof is both a listing of the logical steps and an explanation of the reasoning that went into them.

Obviously you need to know who your audience is. You should assume that the reader is someone, like a fellow student, who has approximately the same background you have. It is important to remember that written proofs have a subjective aspect to them. They are written for a particular audience and how many details you include will depend on who is to read the proof. As you and the rest of the class learn more geometry together, you will share a larger and larger set of common experiences. You can draw on those experiences and assume that your readers will know many of the justifications for steps in the proofs. Later in the course you will be able to leave many of the reasons unstated; this will allow you to concentrate on the essential new ideas in a proof and not obscure them with a lot of detailed information that is already well known to your readers. But don't be too quick to jump ahead to that level. Our aim in this course is to lay out all our assumptions in the axioms and then to base our proofs on those assumptions and no others. Only by being explicit about our reasons for each step can we discipline ourselves to use only those assumptions and not bring in any hidden assumptions that are based on previous experience or on diagrams.

You are encouraged to include in your written proofs more than just a list of the logical steps in the proof. In order to make the proofs more readable, you should also include information for the reader about the structure of the logical argument you are using, what you are assuming, and where the proof is going. Such statements are not strictly necessary from a logical point of view, but they make an enormous difference in the readability of a proof. For now our goal is not so much to prove the theorems as it is to learn to write good proofs, so we will write more than is strictly necessary and not worry about the risk of being pedantic.

The beginning of a proof is marked by the word ***Proof***. It is also a good idea to include an indication that you have reached the end of a proof. Traditionally the end of a proof was indicated with the abbreviation QED, which stands for the Latin phrase *quod erat demonstrandum* (which was to be demonstrated). In this book we mark the end of our proofs with the symbol □.

Like each individual proof, the overall structure of the collection of theorems and proofs in an axiomatic system should be logical and sequential. Within any given proof, it is legitimate to appeal only to the axioms, a theorem that has been previously stated and proved, a definition that has been stated earlier, or to an earlier step in the same proof. The rules of logic that are listed as one possible type of justification for a step in a proof are the rules that are explained in this section. They include such rules as the rules for negating compound statements that were described above and the rules for indirect proofs that will be described below.

There is one last point related to the justification of the steps in a proof that is specific to this course in the foundations of geometry. We intend to build geometry on the real number system. Hence we will base many steps in our proofs on known facts about the real numbers. For example, if we have proved that $x + z = y + z$, we will want to conclude that $x = y$. Technically, this falls under the heading "by previous theorem," but we will usually say something like "by algebra" when we bring in some fact about the algebra of real numbers. Appendix E lists many of the important properties of the real numbers that are assumed in this course. A few of them have names (such as Trichotomy and the Archimedean Property) and those names should be mentioned when the properties are used.

Indirect proof

Indirect proof is one proof form that should be singled out for special consideration because you will find that it is one you will often want to use. The straightforward strategy for proving $P \Rightarrow Q$, called *direct proof*, is to start by assuming that P is true and then to use a series of logical deductions to conclude Q. But the statement $P \Rightarrow Q$ means that Q is true if P is, so the the real purpose of the proof is to rule out the possibility that P is true while Q is false. The indirect strategy is to begin by assuming that P is true *and* that Q is false, and then to show that this leads to a logical contradiction. If it does, then we know that it is impossible for both P and the negation of Q to hold simultaneously and therefore Q must be true whenever P is. This indirect form of proof is called *proof by* *contradiction*. It goes by the official name *reductio ad absurdum*, which we will abbreviate as RAA.

The reason this proof form often works so well is that you have more information with which to work. In a direct argument you assume only the hypothesis P and work from there. In an indirect argument you begin by assuming both the original hypothesis P and also the additional hypothesis *not* Q. You can make use of both assumptions in your proof. In order to help clarify what is going on in an indirect proof, we will give a special name to the additional hypothesis *not* Q; we will call it the *RAA hypothesis* to distinguish it from the standard hypothesis P.

Indirect proofs are often confused with direct proofs of the contrapositive. They are not the same, however, since in a direct proof of the contrapositive we assume only the negation of the conclusion while in an indirect proof we assume and use both the hypothesis and the negation of the conclusion. Suppose we want to prove the theorem P implies Q. A direct proof of the contrapositive would start with *not* Q and conclude *not* P. One way to formulate the argument would be to start by assuming P (the hypothesis) and *not* Q (the RAA hypothesis), then to use the same proof as before to conclude *not* P and finally to end by saying that we must reject the RAA hypothesis because we now have both P and *not* P, an obvious contradiction. While this is logically correct, it is considered to be bad form and sloppy thinking; this way of organizing a proof should, therefore, usually be avoided.[7]

An even worse misuse of proof by contradiction is the following. Suppose again that we want to prove the theorem P implies Q. We assume the hypothesis P. Then we also suppose *not* Q (RAA hypothesis). After that we proceed to prove that P implies Q. At that point in the proof we have both Q and *not* Q. That is a contradiction, so we reject the RAA hypothesis and conclude Q. In this case the structure of an indirect argument has been erected around a direct proof, thus obscuring the real proof. Again this is sloppy thinking. It is an abuse of indirect proof and should *always* be avoided.

Use RAA proofs when they are helpful, but don't misuse them. A proof is often discovered as an indirect proof because we can suppose the conclusion is false and explore the consequences. Once you have found a proof, you should reexamine it to see if the logic can be simplified and the essence of the proof presented more directly.

EXERCISES 2.5

1. Identify the hypothesis and conclusion of each of the following statements.
   (a) If it rains, then I get wet.
   (b) If the sun shines, then we go hiking and biking.
   (c) If $x > 0$, then there exists a y such that $y^2 = 0$.
   (d) If $2x + 1 = 5$, then either $x = 2$ or $x = 3$.
2. State the converse and contrapositive of each of the statements in Exercise 1.
3. Write the negation of each of the statements in Exercise 1.
4. Write each of the following statements in "if..., then..." form.
   (a) It is necessary to score at least 90% on the test in order to receive an A.
   (b) A sufficient condition for passing the test is a score of 50% or higher.
   (c) You fail only if you score less than 50%.
   (d) You succeed whenever you try hard.
5. State the converse and contrapositive of each of the statements in Exercise 4.
6. Write the negation of each of the statements in Exercise 4.
7. Restate each of the following assertions in "if..., then..." form.
   (a) Perpendicular lines must intersect.

[7]We will occasionally find good reason to formulate our proofs this way even though it is generally not good practice.

- **(b)** Any two great circles on $\mathbb{S}^2$ intersect.
- **(c)** Congruent triangles are similar.
- **(d)** Every triangle has angle sum less than or equal to $180°$.

8. Identify the hypothesis and conclusion of each of the following statements.
- **(a)** I can take topology if I pass geometry.
- **(b)** I get wet whenever it rains.
- **(c)** A number is divisible by 4 only if it is even.

9. Restate using quantifiers.
- **(a)** Every triangle has an angle sum of $180°$.
- **(b)** Some triangles have an angle sum of less than $180°$.
- **(c)** Not every triangle has angle sum $180°$.
- **(d)** Any two great circles on $\mathbb{S}^2$ intersect.
- **(e)** If P is a point and ℓ is a line, then there is a line m such that P lies on m and m is perpendicular to ℓ.

10. Negate each of the following statements.
- **(a)** There exists a model for incidence geometry in which the Euclidean Parallel Postulate holds.
- **(b)** In every model for incidence geometry there are exactly seven points.
- **(c)** Every triangle has an angle sum of $180°$.
- **(d)** Every triangle has an angle sum of less than $180°$.
- **(e)** It is hot and humid outside.
- **(f)** My favorite color is red or green.
- **(g)** If the sun shines, then we go hiking.
- **(h)** All geometry students know how to write proofs.

11. Negate each of the three parallel postulates stated in Section 2.3.

12. Construct truth tables that illustrate De Morgan's laws (page 26).

13. Construct a truth table which shows that the conditional statement $H \Rightarrow C$ is logically equivalent to (not H) or C. Then use one of De Morgan's Laws to conclude that not ($H \Rightarrow C$) is logically equivalent to H and (not C).

14. Construct a truth table which shows directly that the negation of $H \Rightarrow C$ is logically equivalent to H and (not C).

15. State the Pythagorean Theorem in "if..., then..." form.

16. Restate each of the three parallel postulates from Section 2.3 in "if..., then..." form.

2.6 SOME THEOREMS FROM INCIDENCE GEOMETRY

We illustrate the lessons of the last section with several theorems and a proof from incidence geometry. The theorems in this section are theorems in incidence geometry, so their proofs are to be based on the three incidence axioms that were stated in §2.2. One of the hardest lessons to be learned in writing the proofs is that we may use only what is actually stated in the axioms, nothing more. Here, again, are the three axioms.

Incidence Axiom 1. *For every pair of distinct points P and Q there exists exactly one line ℓ such that both P and Q lie on ℓ.*

Incidence Axiom 2. *For every line ℓ there exist at least two distinct points P and Q such that both P and Q lie on ℓ.*

Incidence Axiom 3. *There exist three points that do not all lie on any one line.*

The first theorem was already used as an example earlier in the chapter. As explained in the last section, this theorem must be restated before it is ready for a proof. We will adopt the custom of formally restating theorems in if...then... form when necessary.

Definition 2.6.1. Two lines are said to *intersect* if there exists a point that lies on both lines.

Theorem 2.6.2. *Lines that are not parallel intersect in one point.*

Restatement. *If ℓ and m are distinct nonparallel lines, then there exists a unique point P such that P lies on both ℓ and m.*

Proof. Let ℓ and m be two lines such that $\ell \neq m$ and $\ell \nparallel m$ (hypothesis). We must prove two things: first, that there is a point that lies on both ℓ and m and, second, that there is only one such point.

There is a point P such that P lies on both ℓ and m (negation of the definition of parallel). Suppose there exists a second point Q, different from P, such that Q also lies on both ℓ and m (RAA hypothesis). Then ℓ is the unique line such that P and Q lie on ℓ and m is the unique line such that P and Q lie on m (Incidence Axiom 1). Hence $\ell = m$ (definition of unique). But this contradicts the hypothesis that ℓ and m are distinct. Hence we must reject the RAA hypothesis and conclude that no such point Q exists. □

Commentary on the Proof. It is a good idea to begin each proof by restating exactly what the hypotheses give you. (In high school these were called the "givens.") It is also a good idea to state next exactly what it is that you need to prove. This will often amount to nothing more than a restatement of the conclusions of the theorem, but stating them again, in terms of the notation you have used for the hypotheses, helps to focus your thinking and explains to the reader where the proof is going. Once those two preliminaries are out of the way, the real proof begins. In this case there are two things to prove since the conclusion "there exists a unique" means both that there exists one and that there exists only one. The first proof is direct (just a simple invocation of a definition) while the second proof is indirect.

The form of the first proof should be noted. All that we need to do is to look up the definition of parallel and then negate it. When we do that we see that the first conclusion is immediate. This is a useful hint about how to get a proof off the ground: it is often just a matter of going to the definitions of the terms used in the hypotheses and making use of them.

If we were to completely separate the two proofs, we could formulate the second one as a direct proof of the contrapositive (if the point of intersection is not unique, then the lines are not distinct). It is better to formulate the theorem the way we have, however, so that existence and uniqueness can be asserted together.

The reason for each statement is given in parentheses at the end of the sentence. Of course this is only true of those sentences that correspond to logical steps in the proof. The extra, explanatory sentences do not need any justification; their purpose is simply to make the written proof more readable and understandable.

There is one more point regarding the way in which Theorem 2.6.2 is stated that should be clarified. The fact that ℓ and m are lines is not really a logical hypothesis in the theorem, but is just a matter of notation. Thus it might be better to restate the theorem as follows.

Second Restatement of Theorem 2.6.2. *Let ℓ and m be two lines. If ℓ and m are distinct and nonparallel, then there exists a unique point P such that P lies on both ℓ and m.*

One reason it is better to restate the theorem this way is that it is now easy to formulate the converse. Observe that it would not have been as easy if we were working with the original restatement.

Converse to Theorem 2.6.2. *Let ℓ and m be two lines. If there exists a unique point P such that P lies on both ℓ and m, then ℓ and m are distinct and nonparallel.*

Here are several other theorems from incidence geometry. You can practice what you have learned in this chapter by writing proofs for them.

Theorem 2.6.3. *If ℓ is any line, then there exists at least one point P such that P does not lie on ℓ.*

Theorem 2.6.4. *If P is any point, then there are at least two distinct lines ℓ and m such that P lies on both ℓ and m.*

Theorem 2.6.5. *If ℓ is any line, then there exist lines m and n such that ℓ, m, and n are distinct and both m and n intersect ℓ.*

Theorem 2.6.6. *If P is any point, then there exists at least one line ℓ such that P does not lie on ℓ.*

Theorem 2.6.7. *There exist three distinct lines such that no point lies on all three of the lines.*

Theorem 2.6.8. *If P is any point, then there exist points Q and R such that P, Q, and R are noncollinear.*

Theorem 2.6.9. *If P and Q are two points such that $P \neq Q$, then there exists a point R such that P, Q, and R are noncollinear.*

Every theorem has a proper context and that context is an axiomatic system. Thus every theorem has unstated hypotheses, namely that certain axioms are assumed true. For example, the theorems above are theorems in incidence geometry. This means that every one of them includes the unstated hypothesis that the three incidence axioms are assumed true. In the case of Theorem 2.6.7, the unstated hypotheses are the only hypotheses.

One final remark about writing proofs: Except for the gaps we discussed in Chapter 1, Euclid's proofs serve as excellent models for you to follow. Euclid usually includes just the right amount of detail and clearly states his reasons for each step in exactly the way that is advocated in this chapter. He also includes helpful explanations of where the proof is going, so that the reader has a better chance of understanding the big picture. In learning to write good proofs you can do no better than to study Euclid's proofs, especially those from Book I of the *Elements*. As you come to master those proofs you will begin to appreciate them more and more. You will eventually find yourself reading and enjoying not just the proofs themselves, but also Heath's commentary [22] on the proofs. Heath often explains why Euclid did things as he did and also indicates how other geometers have proved the same theorem. Of course Euclid uses language quite differently from the way we do. His theorem statements themselves do not serve as good models of the careful statements that modern standards of rigor demand.

EXERCISES 2.6

1. Prove the converse to Theorem 2.6.2.
2. Prove Theorem 2.6.3.
3. Prove Theorem 2.6.4.
4. Prove Theorem 2.6.5.
5. Prove Theorem 2.6.6.
6. Prove Theorem 2.6.7.
7. Prove Theorem 2.6.8.
8. Prove Theorem 2.6.9.

CHAPTER 3

Axioms for Plane Geometry

This is a course on the *foundations* of geometry and it is in the present chapter that we finally begin to lay the foundations for the geometry we will develop in the course. The preceding chapter explained the necessary background so the reader can now appreciate what it means to organize geometry as an axiomatic system built on a foundation of undefined terms and axioms. The next task is to specify the undefined terms we will use and to state the axioms that spell out what we assume about those terms.

There are choices that must be made: which axioms to use and what assumptions from outside geometry will be allowed into the axioms. Those choices are explored in Appendices B and C and it is recommended that the reader study those appendices in order to better appreciate the thinking behind the selection of axioms that is made in this book. One choice that should be noted here is the fact that the axioms stated in this chapter are based on the real numbers and are stated in the terminology of set theory. The basic information regarding real numbers and set theory that is needed in this and later chapters is reviewed in Appendix E.

The six axioms formulated in this chapter are common to all the geometries studied in the remainder of this book. The geometry that can be done using only these six axioms is called *neutral geometry* because it does not make any assumptions regarding which of the competing postulates of parallelism is true. We will prove our theorems in this neutral setting whenever possible. At the end of the chapter we will restate the three parallel postulates. Much of the remainder of the course will be devoted to an exploration of the logical relationships between the neutral axioms and those parallel postulates.

In addition to statements of the axioms themselves, the chapter includes a number of theorems. The theorems in this chapter are generally foundational in the sense that they spell out the kinds of details that Euclid tacitly assumed but did not prove. This chapter also establishes the basic terminology that will be used throughout the remainder of the course. As a result, the chapter contains a great many definitions. The good news is that most of them will already be familiar to you from your high school geometry course.

The axioms in this chapter are axioms for two-dimensional *plane geometry*. Additional axioms would be needed to describe the geometry of three-dimensional space.

3.1 THE UNDEFINED TERMS AND TWO FUNDAMENTAL AXIOMS

For now there are five undefined terms, namely *point*, *line*, *distance*, *half-plane*, and *angle measure*. In a later chapter we will add *area* to the list, for a total of six undefined terms. In the remainder of this chapter we will specify exactly what is to be assumed about the first five undefined terms. For each of the five terms there is an axiom that spells out what we need to know about that term. There is one additional axiom that specifies how distance and angle measure relate to each other.

The first axiom states rudimentary assumptions about points. The main thing we need to assume about them is that they exist. We will seldom find it necessary to invoke this postulate explicitly, but it plays a key role nonetheless. The geometry in which there are no points and no lines or that in which there is just one point and no lines are not of much interest, but they vacuously satisfy all the remaining neutral axioms. The primary purpose of the Existence Postulate is to rule out those trivial interpretations of the axioms. It is enough to assume the existence of two points because the existence of more points follows from that assumption together with the other postulates stated later in the chapter.

Axiom 3.1.1 (The Existence Postulate). *The collection of all points forms a nonempty set. There is more than one point in that set.*

Definition 3.1.2. The set of all points is called *the plane* and is denoted by $\mathbb{P}$.

The second axiom states our most basic assumptions regarding lines. It is the same as the first axiom of incidence geometry and is also essentially the same as Euclid's first postulate. In his postulate Euclid only asserted that two points determine a line and did not explicitly say that the line is unique, but it is clear from his proofs that he meant it to be unique. As mentioned in the last chapter, the assumption that A and B are *distinct* simply means that they are not the same point.

Axiom 3.1.3 (The Incidence Postulate). *Every line is a set of points. For every pair of distinct points A and B there is exactly one line ℓ such that $A \in \ell$ and $B \in \ell$.*

FIGURE 3.1: Two points determine a unique line

Notation. The notation $\overleftrightarrow{AB}$ is used to indicate the line determined by A and B.

Since a line is a set of points, we can define what it means for a point to lie on a line and it is not necessary to take "lie on" as one of the undefined terms.

Definition 3.1.4. A point P is said to *lie on* line ℓ if $P \in \ell$. The statements "P is incident with ℓ" and "ℓ is incident with P" are also used to indicate the same relationship.

It is also convenient to have a name for the opposite of lie on.

Definition 3.1.5. A point Q is called an *external point* for line ℓ if Q does not lie on ℓ.

This is a good time to repeat a definition that was made earlier.

Definition 3.1.6. Two lines ℓ and m are said to be *parallel*, written $\ell \parallel m$, if there is no point P such that P lies on both ℓ and m (i.e., if $\ell \cap m = \emptyset$).

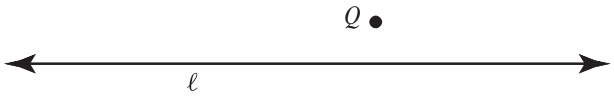

FIGURE 3.2: The point Q is an external point for line ℓ

Be sure to notice that, according to this definition, a line is not parallel to itself. This may conflict with the way the word *parallel* was used in your high school geometry course. We are now ready for our first theorem.

Theorem 3.1.7. *If ℓ and m are two distinct, nonparallel lines, then there exists exactly one point P such that P lies on both ℓ and m.*

Proof. This is a restatement of the first theorem of incidence geometry, Theorem 2.6.2. The proof given earlier is still valid since it was based on just the Incidence Postulate. □

The theorem can be restated as a type of trichotomy for pairs of lines. If ℓ and m are two lines, then exactly one of the following conditions will hold: Either $\ell = m$, $\ell \parallel m$, or $\ell \cap m$ consists of precisely one point.

3.2 DISTANCE AND THE RULER POSTULATE

The third axiom spells out what is assumed regarding the third undefined term, *distance*.

Axiom 3.2.1 (The Ruler Postulate). *For every pair of points P and Q there exists a real number PQ, called the distance from P to Q. For each line ℓ there is a one-to-one correspondence from ℓ to $\mathbb{R}$ such that if P and Q are points on the line that correspond to the real numbers x and y, respectively, then $PQ = |x - y|$.*

There is a lot packed into this statement: the Ruler Postulate not only specifies the basic properties of distance measurement, but it also implies that it is possible to introduce (real number) coordinates on a line and that a line is continuous (with no gaps). Since this postulate is so crucial to our development of geometry, we will carefully unpack its statement in this section and explore its various technical implications. We begin with several fundamental definitions.

Definition 3.2.2. Three points A, B, and C are *collinear* if there exists one line ℓ such that A, B, and C all lie on ℓ. The points are *noncollinear* otherwise.

Definition 3.2.3. Let A, B, and C be three distinct points. The point C is *between* A and B, written $A * C * B$, if A, B, and C are collinear and $AC + CB = AB$.

We will see in the next chapter that the condition $AC + CB = AB$ implies that A, B, and C are collinear, so we will be able to drop the collinearity assumption from the definition of between. But that proof relies on axioms we have not yet stated. Without additional assumptions both conditions are necessary; see Exercises 3.2.5 and 3.2.6.

Definition 3.2.4. Define the *segment* $\overline{AB}$ by

$$\overline{AB} = \{A, B\} \cup \{P \mid A * P * B\}$$

and the *ray* $\overrightarrow{AB}$ by

$$\overrightarrow{AB} = \overline{AB} \cup \{P \mid A * B * P\}.$$

The segment $\overline{AB}$ consists of the two points A and B together with all the points between A and B. The ray $\overrightarrow{AB}$ contains the segment $\overline{AB}$ together with all the points that are "beyond" B in the sense that B is between A and the point. The notation used is consistent with the notation $\overleftrightarrow{AB}$ for the line determined by A and B. In each case the notation suggests the shape of the set. (See Figure 3.3.)

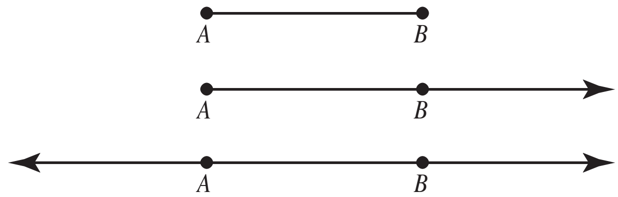

FIGURE 3.3: Segment, ray, line

Definition 3.2.5. The *length* of segment $\overline{AB}$ is AB, the distance from A to B. Two segments $\overline{AB}$ and $\overline{CD}$ are said to be *congruent*, written $\overline{AB} \cong \overline{CD}$, if they have the same length.

Definition 3.2.6. The points A and B are the *endpoints* of the segment $\overline{AB}$; all other points of $\overline{AB}$ are *interior points*. The point A is the *endpoint* of the ray $\overrightarrow{AB}$.

The Ruler Postulate really asserts the existence of two different kinds of real-valued functions. One function measures distances between points and the other assigns coordinates to points on a line in a way that is consistent with the overall distance-measuring function. The next theorem lists the basic properties of the function that measures distance. The proof illustrates the way in which the postulate allows us to bring facts about the algebra of real numbers to bear on geometry.

Theorem 3.2.7. *If P and Q are any two points, then*

1. $PQ = QP$,
2. $PQ \geqslant 0$, *and*
3. $PQ = 0$ *if and only if* $P = Q$.

Proof. Let P and Q be two points (hypothesis). We will first show that there is a line ℓ such that both P and Q lie on ℓ. Either $P = Q$ or $P \neq Q$ and we consider each case separately. Suppose $P \neq Q$. Then there is exactly one line $\ell = \overleftrightarrow{PQ}$ such that P and Q lie on ℓ (Incidence Postulate). If $P = Q$, there must be another point $R \neq P$ (Existence Postulate). In that case we take ℓ to be the unique line such that P and R lie on ℓ (Incidence Postulate). In either case ℓ is a line such that both P and Q lie on ℓ.

There exists a one-to-one correspondence from ℓ to $\mathbb{R}$ having the properties specified in the Ruler Postulate. In particular, P and Q correspond to real numbers x and y, respectively, with $PQ = |x - y|$ and $QP = |y - x|$ (Ruler Postulate). But $|x - y| = |y - x|$ (algebra), so $PQ = QP$. Also $|x - y| \geqslant 0$ (algebra), so $PQ \geqslant 0$. This proves conclusions (1) and (2) of the theorem.

Suppose $PQ = 0$ (hypothesis). Then $|x - y| = 0$, so $x = y$ (algebra). Therefore, $P = Q$ (the correspondence is one-to-one). Finally, if $P = Q$, then $x = y$ (the correspondence is a function), so $PQ = |x - x| = 0$. This completes the proof of both parts of (3). □

Corollary 3.2.8. *$A * C * B$ if and only if $B * C * A$.*

Proof. Let A, B, and C be three points such that $C \in \overleftrightarrow{AB}$ (hypothesis). If $A * C * B$, then $AC + CB = AB$ (definition). Since $AB = BA$, $AC = CA$, and $CB = BC$, it is also the case that $BC + CA = BA$. Therefore, $B * C * A$. The proof of the converse is similar. □

Any function that has the properties spelled out in Theorem 3.2.7 can be used to measure distances. Such a function is called a metric.

Definition 3.2.9. A *metric* is a function $D : \mathbb{P} \times \mathbb{P} \rightarrow \mathbb{R}$ such that

1. $D(P, Q) = D(Q, P)$ for every P and Q,
2. $D(P, Q) \geqslant 0$ for every P and Q, and
3. $D(P, Q) = 0$ if and only if $P = Q$.

A version of the triangle inequality

$$D(P, Q) \leqslant D(P, R) + D(R, Q)$$

is often included as part of the definition of metric, but we do not include it in the definition because we will prove the triangle inequality as a theorem in the next chapter.[1]

The familiar distance formula from calculus and high school geometry is an example of a metric.

EXAMPLE 3.2.10 The Euclidean metric

Define the distance between points (x_1, y_1) and (x_2, y_2) in the Cartesian plane by

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

This metric is called the *Euclidean metric*. The verification that d satisfies the conditions in the definition of metric is left as an exercise (Exercise 2). ■

There are other ways in which to measure distances in the Cartesian plane. Here is an example that will be important later in the chapter.

EXAMPLE 3.2.11 The taxicab metric

Define the distance between points (x_1, y_1) and (x_2, y_2) in the Cartesian plane by

$$\rho((x_1, y_1), (x_2, y_2)) = |x_2 - x_1| + |y_2 - y_1|.$$

This metric is called the *taxicab metric*. The verification that ρ satisfies the conditions in the definition of metric is left as an exercise (Exercise 3). ■

The name *taxicab* is given to the second metric because the distance from (x_1, y_1) to (x_2, y_2) is measured by traveling along a line parallel to the x-axis and then along one parallel to the y-axis in much the same way that a taxicab in Manhattan must follow a rectangular grid of streets and cannot travel directly from point A to point B. There is also a natural way to measure distances between points on the sphere.

[1] What we are calling a metric should more properly be called a *semimetric*. However, all the examples of metrics we will consider satisfy the triangle inequality and are metrics according to the standard definition. We omit the triangle inequality from the definition only because it is an unnecessary assumption in this context.

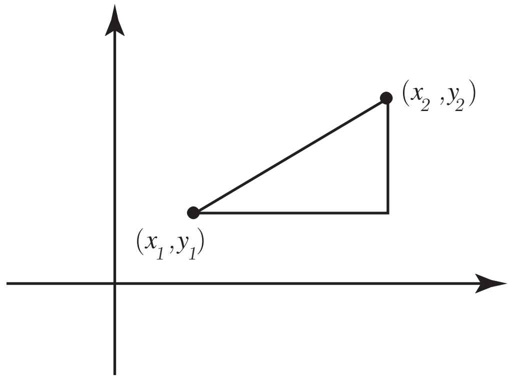

FIGURE 3.4: The Euclidean distance from (x_1, y_1) to (x_2, y_2) is the length of the "straight" line segment while the taxicab distance is the sum of the horizontal and vertical lengths

EXAMPLE 3.2.12 The spherical metric

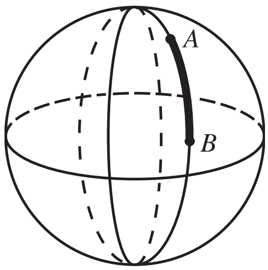

FIGURE 3.5: The spherical distance from A to B is the length of the shorter arc of the great circle

Let A and B be two points on the sphere $\mathbb{S}^2$ (Example 2.2.9). Define $s(A, B)$, the *spherical distance from A to B*, to be the length of the shortest arc of a great circle containing A and B. More specifically, if A and B are antipodal points, define $s(A, B) = \pi$. If A and B are not antipodal points, then they lie on a unique great circle and they are the endpoints of two subarcs of that great circle. The distance from A to B is defined to be the length of the shorter of the two subarcs. The verification that s is a metric is left as an exercise (Exercise 4). In this metric, no distance is greater than π. ■

We now turn our attention to the second part of the Ruler Postulate. The assertion that there is a one-to-one correspondence from ℓ to $\mathbb{R}$ means that there is a function $f : \ell \to \mathbb{R}$ such that f is both one-to-one and onto. The postulate says that for any two points P and Q that lie on ℓ, $PQ = |f(P) - f(Q)|$. Such functions are useful because they facilitate the transition from geometry to algebra.

Definition 3.2.13. Let ℓ be a line. A one-to-one correspondence $f : \ell \to \mathbb{R}$ such that $PQ = |f(P) - f(Q)|$ for every P and Q on ℓ is called a *coordinate function* for the line ℓ and the number $f(P)$ is called the *coordinate of the point P*.

It is possible to construct coordinate functions associated with each of the metrics we have defined for the Cartesian plane.

EXAMPLE 3.2.14 Coordinate functions in the Euclidean metric

If ℓ is a nonvertical line in the Cartesian plane, then ℓ has an equation of the form $y = mx + b$. In that case we define $f : \ell \to \mathbb{R}$ by $f(x, y) = x\sqrt{1 + m^2}$. If ℓ is a vertical line, then it has an equation of the form $x = a$. In that case we define f by $f(a, y) = y$. The verification that these functions are coordinate functions in the Euclidean metric is left as an exercise (Exercise 9). ■

EXAMPLE 3.2.15 Coordinate functions in the taxicab metric

If ℓ is a nonvertical line in the Cartesian plane with equation $y = mx + b$, define $f : \ell \to \mathbb{R}$ by $f(x, y) = x(1 + |m|)$. If ℓ is a vertical line with equation $x = a$, define f by $f(a, y) = y$. The verification that these functions are coordinate functions in the taxicab metric is left as an exercise (Exercise 10). ■

It should be emphasized that the metrics and coordinate functions we have defined are merely examples and we are not restricting ourselves to any one interpretation of the undefined terms such as distance. While we describe these examples to help clarify the meaning of the Ruler Postulate, our approach to geometry is axiomatic and we remain open to any interpretation of the undefined terms that could serve as a model for the axioms.

There are no coordinate functions for lines in spherical geometry because the distance between points on a line (= great circle) on $\mathbb{S}^2$ has a finite upper bound while the existence of a coordinate function for a line implies that there are points on the line with arbitrarily large distances between them. We will see later in the course that it is possible to construct a metric and corresponding coordinate functions for lines in the Klein disk, but the definitions of those functions requires substantial technical preparation.

Using the definition of coordinate function, the second part of the Ruler Postulate can be restated quite simply: *For every line there is a coordinate function.* The postulate asserts only existence, not uniqueness; each line will have many different coordinate functions associated with it. Given two points on line ℓ, it is often convenient to arrange that the first is assigned the coordinate 0 and and the second has positive coordinate. The next theorem assures us that this can always be done. It is the first of many examples in this book of results that are proved as theorems but, for historical reasons, are referred to as postulates. This particular statement originated in the list of axioms introduces by the School Mathematics Study Group (SMSG) in the 1960s [40]. It is still listed as an axiom in most high school textbooks, but we prefer to prove it as a consequence of the Ruler Postulate. Two technical parts of the proof will be left as exercises.

Theorem 3.2.16 (The Ruler Placement Postulate). *For every pair of distinct points P and Q, there is a coordinate function $f : \overleftrightarrow{PQ} \to \mathbb{R}$ such that $f(P) = 0$ and $f(Q) > 0$.*

Proof. Fix two distinct points P and Q (hypothesis) and let $\ell = \overleftrightarrow{PQ}$ (Incidence Postulate). There exists a coordinate function $g : \ell \to \mathbb{R}$ (Ruler Postulate). Define $c = -g(P)$ and define $h : \ell \to \mathbb{R}$ by $h(X) = g(X) + c$ for every $X \in \ell$. Then h is a coordinate function (Exercise 12(b)). Note that $h(P) = 0$. Now $h(Q) \neq 0$ (h is 1-1) so either $h(Q) > 0$ or $h(Q) < 0$ (trichotomy). If $h(Q) > 0$, then $f = h$ satisfies the conclusion of the Ruler Placement Postulate. If $h(Q) < 0$, define $f : \ell \to \mathbb{R}$ by $f(X) = -h(X)$. Then f is a coordinate function (Exercise 12(a)); furthermore, $f(P) = 0$ and $f(Q) > 0$, so the proof is complete. □

This is a good time to reflect on the definition of betweenness for points that was stated at the beginning of the section and how it relates to coordinate functions. It would be natural to use the ordering of the real numbers to define betweenness; we could define B to be between A and C if the coordinate of B is between the coordinates of A and C in the natural ordering of numbers on the real line. This is, in fact, the way betweenness is defined in high school geometry textbooks. (See [45], page 46, for example.) The problem with this definition is that it then becomes necessary to prove that betweenness is well defined, that it does not depend on which coordinate function is used for the line containing the three points. In order to avoid that minor technicality, we have followed the majority of college-level geometry books in defining betweenness using the equation $AB + BC = AC$. Either way it is necessary to prove a theorem asserting that the two possible definitions are equivalent. A complete proof of the following betweenness theorem is included in the margin of the teachers' edition of [45] (see page 50).

Theorem 3.2.17 (Betweenness Theorem for Points). *Let ℓ be a line; let A, B, and C be three distinct points that all lie on ℓ; and let $f : \ell \to \mathbb{R}$ be a coordinate function for ℓ. The point C is between A and B if and only if either $f(A) < f(C) < f(B)$ or $f(A) > f(C) > f(B)$.*

Proof. Let ℓ be a line, let A, B, and C be three distinct points that all lie on ℓ, and let $f : \ell \to \mathbb{R}$ be a coordinate function for ℓ (hypothesis). If $f(A) < f(C) < f(B)$, then

$$|f(C) - f(A)| + |f(B) - f(C)| = |f(B) - f(A)|$$

(algebra), so C is between A and B (definition of between). In a similar way it follows that C is between A and B if $f(A) > f(C) > f(B)$. This completes the proof of one half of the theorem.

Now suppose C is between A and B (hypothesis). Then

$$|f(C) - f(A)| + |f(B) - f(C)| = |f(B) - f(A)|$$

(definition of between). It is also the case that $(f(C) - f(A)) + (f(B) - f(C)) = (f(B) - f(A))$ (no absolute values), so $f(C) - f(A)$ and $f(B) - f(C)$ have the same sign; that is, either both are positive or both are negative (algebra).[2] In case both are positive, $f(A) < f(C) < f(B)$ (algebra). In case both are negative, $f(A) > f(C) > f(B)$ (algebra). This completes the proof. □

Corollary 3.2.18. *Let A, B, and C be three points such that B lies on $\overrightarrow{AC}$. Then $A * B * C$ if and only if $AB < AC$.*

Corollary 3.2.19. *If A, B, and C are three distinct collinear points, then exactly one of them lies between the other two.*

Proof. Under any coordinate function, the three distinct points correspond to three distinct real numbers x, y, and z (the coordinate function is one-to-one). The numbers x, y, and z can be ordered from smallest to largest and, by Theorem 3.2.17, this uniquely determines which point is between the others on the line. □

Corollary 3.2.20. *Let A and B be two distinct points. If f is a coordinate function for $\ell = \overleftrightarrow{AB}$ such that $f(A) = 0$ and $f(B) > 0$, then $\overrightarrow{AB} = \{P \in \ell \mid f(P) \geqslant 0\}$.*

Proof. Exercise 15. □

Definition 3.2.21. Let A and B be two distinct points. The point M is called a *midpoint* of $\overline{AB}$ if M is between A and B and $AM = MB$.

[2]The fact from high school algebra being used is this: If x and y are two nonzero real numbers such that $|x| + |y| = |x + y|$, then either both x and y are positive or both x and y are negative (Exercise 14).

The midpoint of $\overline{AB}$ is between A and B, so $AM + MB = AB$. Simple algebra shows that $AM = (1/2)AB = MB$.

FIGURE 3.6: M is the midpoint of $\overline{AB}$

The next theorem asserts that midpoints exist and are unique. For us the proof is an exercise in the use of coordinate functions. Euclid also proved that midpoints exist (Proposition I.10), but for Euclid this meant proving that the midpoint can be constructed using compass and straightedge.

Theorem 3.2.22 (Existence and Uniqueness of Midpoints). *If A and B are distinct points, then there exists a unique point M such that M is the midpoint of $\overline{AB}$.*

Proof. Exercise 17. □

You might wonder why we bother to prove that a midpoint exists when we did not stop to prove that other objects we defined earlier (such as segments, rays, etc.) exist. It is easy to define something like a midpoint, but the act of making the definition does not by itself guarantee that there is anything that satisfies the definition. For example, we could define "the first point in the interior of $\overline{AB}$" to be the point F in the interior of $\overline{AB}$ such that $F * C * B$ for every point $C \neq F$ in the interior of $\overline{AB}$. Even though this reads like a reasonable definition, every calculus student knows that no such point exists. For this reason it is a good idea to prove that something like a midpoint exists. At the same time this should not be taken to an unreasonable extreme; it is not necessary, for example, to prove that the segment determined by two points A and B exists. The segment was defined to be the set of points that satisfy a certain condition. This set exists whether or not there are any points that satisfy the condition. (Of course it would be easy to use the Ruler Postulate to prove that there are lots of points in $\overline{AB}$.) Similar comments apply to other objects we have defined.

The technical complication of some of the definitions and theorems we have discussed in this section may give the impression that the Ruler Postulate is difficult to apply because we always have to deal with coordinate functions. In fact that is not the case at all. We have explored the ramifications of the statement to ensure a deep understanding of what it implies, but the more technical aspects of the Ruler Postulate are rarely needed. The most common application of the Ruler Postulate is to locate a point at the correct distance from the endpoint of a ray. The next result will be used frequently in the remainder of the course and accounts for the vast majority of the invocations of the Ruler Postulate. The statement is intuitive and clear, so it can be thought of as an axiom.

Theorem 3.2.23 (Point Construction Postulate). *If A and B are distinct points and d is any nonnegative real number, then there exists a unique point C such that C lies on $\overrightarrow{AB}$ and $AC = d$.*

Proof. Let A and B be distinct points and let $d \geq 0$ be a real number (hypothesis). We must prove two things: first, there is a point C on $\overrightarrow{AB}$ such that $AC = d$ and, second, there is only one such point.

There is a coordinate function $f : \overleftrightarrow{AB} \to \mathbb{R}$ such that $f(A) = 0$ and $f(B) > 0$ (Ruler Placement Postulate). There exists a point $C \in \overleftrightarrow{AB}$ such that $f(C) = d$ (f is onto) and $C \in \overrightarrow{AB}$ (Corollary 3.2.20). Now $AC = |f(A) - f(C)| = |0 - d| = d$ (f is a coordinate function), so existence is proved.

Assume C' is a point on $\overrightarrow{AB}$ such that $AC' = d$. Then $f(C') \geq 0$ (Corollary 3.2.20) and $|f(C')| = |0 - f(C')| = |f(A) - f(C')| = AC' = d$ (f is a coordinate function).

Hence $f(C') = f(C)$. But f is one-to-one (definition of coordinate function), so $C' = C$. This completes the proof of uniqueness. □

The final aspect of the Ruler Postulate that we wish to examine is the fact that it specifies that a coordinate function is a one-to-one correspondence between points on a line and the *real* numbers. The fact that the real numbers are used, rather than some other system of numbers, ensures that there are no "holes" or "gaps" in a line or a circle. There is nothing in the other postulates we have stated that ensures this, but it is a property we need.

EXAMPLE 3.2.24 The rational plane

Interpret *point* to mean an ordered pair of rational numbers. The collection of such ordered pairs is called the *rational plane* and a point in the rational plane is called a *rational point*. Since every rational number is also a real number, the rational plane is a subset of the Cartesian plane. We interpret *line* to mean all the rational points on a Cartesian line that has rational slope and intercept. In other words, a line in the rational plane is a set of the form

$$\ell = \{(r, s) \mid r \text{ and } s \text{ are rational numbers and } ar + bs + c = 0\}$$

for some fixed rational numbers a, b, and c with a and b not both 0. We measure distances in the rational plane using the Euclidean metric.

The Euclidean metric is a metric on the rational plane, so the rational plane satisfies the first part of the Ruler Postulate. It also satisfies the Existence and Incidence Postulates. The rational plane does not satisfy the second part of the Ruler Postulate because there is no one-to-one correspondence between the points on a line in the rational plane and the real numbers. It is easy to see that there is a one-to-one correspondence between points on a rational line and the rational numbers. But a theorem from set theory[3] asserts that there is no one-to-one correspondence from $\mathbb{Q}$ to $\mathbb{R}$, so there can be no one-to-one correspondence between the points on a rational line and the real numbers. ■

The rational plane satisfies all five of Euclid's postulates. However, the proof of Euclid's very first proposition, which we examined in Chapter 1, breaks down in the rational plane.

Definition 3.2.25. Given a point O and a positive real number r, the *circle with center O and radius r* is defined to be the set of all points P such that the distance from O to P is r.

EXAMPLE 3.2.26 Circles in the rational plane

Consider the rational points $A = (0, 0)$ and $B = (2, 0)$. The circles of radius 2 centered at A and B do not intersect in the rational plane. There are two points in the Cartesian plane at which the corresponding circles intersect, namely $(1, \pm\sqrt{3})$, but those points are not points in the rational plane. ■

The rational plane shows that Euclid was indeed using unstated hypotheses in his proofs. Since the rational plane satisfies all of Euclid's postulates, Euclid's implicit assertion in the proof of Proposition 1 that the two circles he constructs must intersect is based on assumptions that are not made explicit in the postulates. According to modern standards of rigor, this means that there is a gap in Euclid's proof. Our assumption that

[3] In 1891 Georg Cantor used his famous diagonal argument to prove that there is no one-to-one correspondence from the rational numbers to the real numbers.

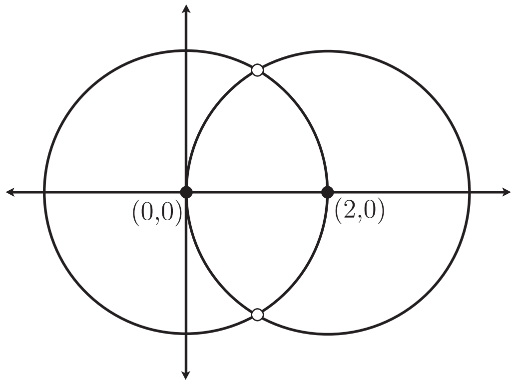

FIGURE 3.7: Two circles in the rational plane

coordinate functions use real numbers will allow us to fill that gap. Specifically, we will prove that if one circle contains a point that is inside a second circle and another point that is outside the second circle, then the two circles must intersect. Even using the powerful axioms in this chapter the proof is not easy, however, and we will not attempt that particular proof until Chapter 8.

EXERCISES 3.2

1. Prove: If ℓ and m are two lines, the number of points in $\ell \cap m$ is either 0, 1, or ∞.

2. Show that the Euclidean metric defined in Example 3.2.10 is a metric (i.e., verify that the function d satisfies the three conditions in the definition of metric on page 39).

3. Show that the taxicab metric defined in Example 3.2.11 is a metric (i.e., verify that the function ρ satisfies the three conditions in the definition of metric on page 39).

4. Show that the spherical metric defined in Example 3.2.12 is a metric (i.e., verify that the function s satisfies the three conditions in the definition of metric on page 39).

5. Betweenness in taxicab geometry. Let $A = (0, 0)$, $B = (1, 0)$ and $C = (1, 1)$ and let ρ denote the taxicab metric.

(a) Find all points P such that $\rho(A, P) + \rho(P, B) = \rho(A, B)$. Draw a sketch in the Cartesian plane.

(b) Find all points P such that $\rho(A, P) + \rho(P, C) = \rho(A, C)$. Draw a sketch in the Cartesian plane.

6. Betweenness on the sphere. Let A, B, and C be points on the sphere $\mathbb{S}^2$. Define C to be *between* A and B if A, B, and C are collinear and $s(A, C) + s(C, B) = s(A, B)$, where s is the metric defined in Example 3.2.12. Also define *segment* in the usual way.

(a) Find all points that are between A and C in case A and C are nonantipodal points. Sketch the segment from A to C.

(b) Find all points that are between A and C in case A and C are antipodal points. Sketch the segment from A to C.

7. Find all points (x, y) in $\mathbb{R}^2$ such that $\rho((0, 0), (x, y)) = 1$, where ρ is the taxicab metric. Draw a sketch in the Cartesian plane. (This shape might be called a "circle" in the taxicab metric.)

8. The square metric. Define the distance between two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ by $D((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$.[4]

(a) Verify that D is a metric.

(b) Find all points (x, y) in $\mathbb{R}^2$ such that $D((0, 0), (x, y)) = 1$. Draw a sketch in the Cartesian plane. (This should explain the name *square metric*.)

[4] $\max\{a, b\}$ denotes the larger of the two real numbers a and b.

9. Verify that the functions defined in Example 3.2.14 are coordinate functions.
10. Verify that the functions defined in Example 3.2.15 are coordinate functions.
11. Let ℓ be a line in the Cartesian plane $\mathbb{R}^2$. Find a function $f : \ell \to \mathbb{R}$ that is a coordinate function for ℓ in the square metric. (See Exercise 8 for definition of square metric.)
12. Assume that $f : \ell \to \mathbb{R}$ is a coordinate function for ℓ.
  (a) Prove that $-f$ is also a coordinate function for ℓ.
  (b) Prove that $g : \ell \to \mathbb{R}$ defined by $g(P) = f(P) + c$ for some constant c is also a coordinate function for ℓ.
  (c) Prove that if $h : \ell \to \mathbb{R}$ is any coordinate function for ℓ then there must exist a constant c such that either $h(P) = f(P) + c$ or $h(P) = -f(P) + c$.
13. Let ℓ be a line and let $f : \ell \to \mathbb{R}$ be a function such that $PQ = |f(P) - f(Q)|$ for every $P, Q \in \ell$. Prove that f is a coordinate function for ℓ.
14. Prove the following fact from high school algebra that was needed in the proof of Theorem 3.2.17: If x and y are two nonzero real numbers such that $|x| + |y| = |x + y|$, then either both x and y are positive or both x and y are negative.
15. Prove Corollary 3.2.20.
16. Prove that if $C \in \overrightarrow{AB}$ and $C \neq A$, then $\overrightarrow{AB} = \overrightarrow{AC}$.
17. Prove existence and uniqueness of midpoints (Theorem 3.2.22).
18. (Segment Construction Theorem) Prove the following theorem. If $\overline{AB}$ is a segment and $\overrightarrow{CD}$ is a ray, then there is a unique point E on $\overrightarrow{CD}$ such that $\overline{AB} \cong \overline{CE}$.
19. (Segment Addition Theorem) Prove the following theorem. If $A * B * C$, $D * E * F$, $\overline{AB} \cong \overline{DE}$, and $\overline{BC} \cong \overline{EF}$, then $\overline{AC} \cong \overline{DF}$.
20. (Segment Subtraction Theorem) Prove the following theorem. If $A * B * C$, $D * E * F$, $\overline{AB} \cong \overline{DE}$, and $\overline{AC} \cong \overline{DF}$, then $\overline{BC} \cong \overline{EF}$.
21. Let A and B be two distinct points. Prove that $\overline{AB} = \overline{BA}$.
22. Prove that the endpoints of a segment are well defined (i.e., if $\overline{AB} = \overline{CD}$, then either $A = C$ and $B = D$ or $A = D$ and $B = C$).
23. Let A and B be two distinct points. Prove each of the following equalities.
  (a) $\overrightarrow{AB} \cup \overrightarrow{BA} = \overleftrightarrow{AB}$
  (b) $\overrightarrow{AB} \cap \overrightarrow{BA} = \overline{AB}$
24. Let A, B, and C be three collinear points such that $A * B * C$. Prove each of the following set equalities.
  (a) $\overrightarrow{BA} \cup \overrightarrow{BC} = \overleftrightarrow{AC}$
  (b) $\overrightarrow{BA} \cap \overrightarrow{BC} = \{B\}$
  (c) $\overline{AB} \cup \overline{BC} = \overline{AC}$
  (d) $\overline{AB} \cap \overline{BC} = \{B\}$
  (e) $\overrightarrow{AB} = \overrightarrow{AC}$

3.3 PLANE SEPARATION

The fourth axiom explains how a line divides the plane into two half-planes. Among other things, this axiom allows us to define angle—one of the most basic objects of study in geometry.

Definition 3.3.1. A set of points S is said to be a *convex set* if for every pair of points A and B in S, the entire segment $\overline{AB}$ is contained in S.

Axiom 3.3.2 (The Plane Separation Postulate). *For every line ℓ, the points that do not lie on ℓ form two disjoint, nonempty sets H_1 and H_2, called half-planes bounded by ℓ, such that the following conditions are satisfied:*

1. *Each of H_1 and H_2 is convex.*
2. *If $P \in H_1$ and $Q \in H_2$, then $\overline{PQ}$ intersects ℓ.*

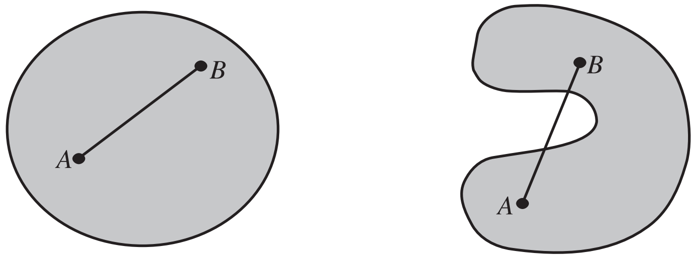

FIGURE 3.8: Convex, not convex

Some of what the postulate asserts can be more succinctly stated using set theoretic notation. Here, in symbols, is what the postulate says about H_1 and H_2.

- $H_1 \bigcup H_2 = \mathbb{P} \setminus \ell$.
- $H_1 \bigcap H_2 = \emptyset$.
- $H_1 \neq \emptyset$ and $H_2 \neq \emptyset$.
- If $A \in H_1$ and $B \in H_1$, then $\overline{AB} \subseteq H_1$ and $\overline{AB} \bigcap \ell = \emptyset$.
- If $A \in H_2$ and $B \in H_2$, then $\overline{AB} \subseteq H_2$ and $\overline{AB} \bigcap \ell = \emptyset$.
- If $A \in H_1$ and $B \in H_2$, then $\overline{AB} \bigcap \ell \neq \emptyset$.

Notation. Let ℓ be a line and let A be an external point. We use the notation H_A to denote the half-plane bounded by ℓ that contains A.

Definition 3.3.3. Let ℓ be a line, let H_1 and H_2 be the two half-planes bounded by ℓ, and let A and B be two external points. We say that A and B are *on the same side of* ℓ if they are both in H_1 or both in H_2. The points A and B are *on opposite sides of* ℓ if one is in H_1 and the other is in H_2.

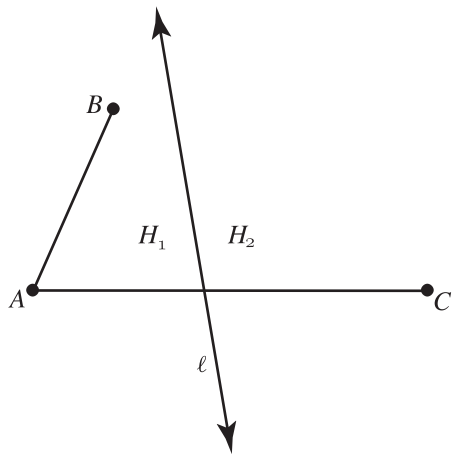

FIGURE 3.9: A and B are on the same side of ℓ; A and C are on opposite sides of ℓ

The separation properties of a line can be restated in the terms just defined.

Proposition 3.3.4. *Let ℓ be a line and let A and B be points that do not lie on ℓ. The points A and B are on the same side of ℓ if and only if $\overline{AB} \bigcap \ell = \emptyset$. The points A and B are on opposite sides of ℓ if and only if $\overline{AB} \bigcap \ell \neq \emptyset$.*

Definition 3.3.5. Two rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ having the same endpoint are *opposite* rays if the two rays are unequal but $\overleftrightarrow{AB} = \overleftrightarrow{AC}$. Otherwise they are *nonopposite*.

Another way to state the definition is this: $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are opposite rays if $B * A * C$.

FIGURE 3.10: Opposite rays

Definition 3.3.6. An *angle* is the union of two nonopposite rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ sharing the same endpoint. The angle is denoted by either $\angle BAC$ or $\angle CAB$. The point A is called the *vertex* of the angle and the rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are called the *sides* of the angle.

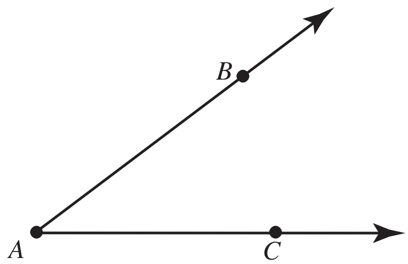

FIGURE 3.11: An angle

Notice that the angle $\angle BAC$ is the same as the angle $\angle CAB$. We have defined angle in such a way that there is no angle determined by opposite rays. This may differ from what you learned in high school, where you probably worked with straight angles. One reason we define angle the way we do is so that we can go on to make the next definition. The definition of interior of an angle would not make sense for a straight angle (or for larger angles).

Definition 3.3.7. Let A, B, and C be three points such that the rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are nonopposite. The *interior* of angle $\angle BAC$ is defined as follows. If $\overrightarrow{AB} \neq \overrightarrow{AC}$, then the interior of $\angle BAC$ is defined to be the intersection of the half-plane H_B determined by B and $\overleftrightarrow{AC}$ and the half-plane H_C determined by C and $\overleftrightarrow{AB}$ (i.e., the interior of $\angle BAC$ is the set of points $H_B \cap H_C$). If $\overrightarrow{AB} = \overrightarrow{AC}$, then the interior of $\angle BAC$ is defined to be the empty set.

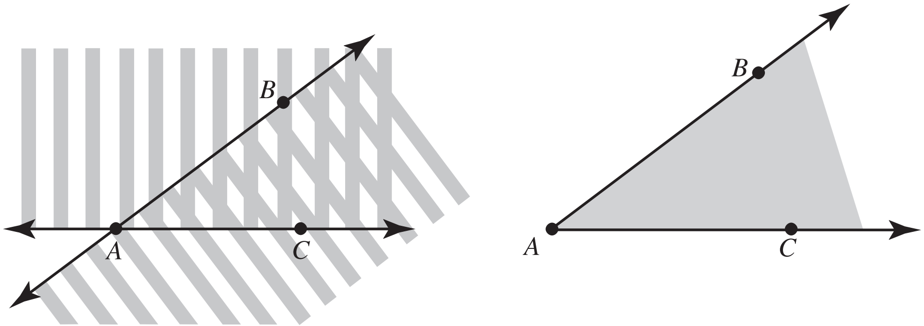

FIGURE 3.12: The two half-planes determined by an angle and the interior of the angle

The interior of any angle is a convex set; see Exercise 1. Since a half-plane consists of all points on the same side of a given line, another way to state the definition of angle interior is this: A point P is in the interior of $\angle BAC$ if P is on the same side of $\overleftrightarrow{AB}$ as C and on the same side of $\overleftrightarrow{AC}$ as B. Among other things, the concept of angle interior allows us to define betweenness for rays.

Definition 3.3.8. Ray $\overrightarrow{AD}$ is *between* rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ if D is in the interior of $\angle BAC$.

There is a minor technical issue related to the last two definitions that should be addressed. Since the angle $\angle BAC$ is defined to be the union of the two rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$, we need to be sure that the interior of the angle is well defined in the sense that it depends only on the two rays and not on the particular points B and C that are used to describe them. (The vertex A is uniquely determined by the angle—see Exercise 6.) In a similar way, whether $\overrightarrow{AD}$ is between rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ should depend only on the ray $\overrightarrow{AD}$ and not on the particular point D that is used to describe it. Assurance on both points is provided by the following rather technical theorem. In particular, D is in the interior of $\angle BAC$ if and only if D' is in the interior of $\angle BAC$ for every $D' \neq A$ on $\overrightarrow{AD}$ (see Figure 3.13).

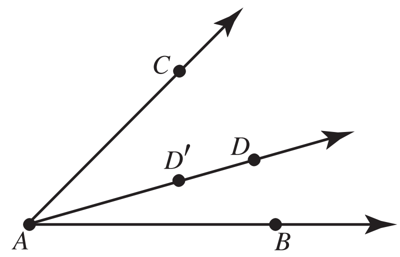

FIGURE 3.13: Betweenness for rays

Theorem 3.3.9 (The Ray Theorem). *Let ℓ be a line, A a point on ℓ, and B an external point for ℓ. If C is a point on $\overrightarrow{AB}$ and $C \neq A$, then B and C are on the same side of ℓ.*

Proof. Let ℓ be a line, let A be a point on ℓ, let B be an external point for ℓ, and let C be a point of $\overrightarrow{AB}$ that is different from A (hypothesis). We must prove that $\overline{BC} \cap \ell = \emptyset$ (definition of same side). There are two cases: either $A * C * B$ or $A * B * C$ (definition of ray).

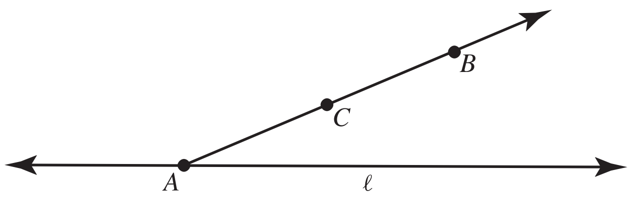

FIGURE 3.14: B and C are on the same side of ℓ (Theorem 3.3.9)

Assume, first that $A * C * B$. Then A is not between B and C (Corollary 3.2.19) so A is not on the segment $\overline{BC}$. The lines ℓ and $\overleftrightarrow{AB}$ have only one point in common (Theorem 3.1.7) and that point must be A, which is not on $\overline{BC}$ (previous statement), so $\overline{BC} \cap \ell = \emptyset$. This completes the proof of the first case.

Now assume $A * B * C$. Then B and C are on the same side of ℓ by the same argument as in the previous paragraph, but with the roles of B and C reversed. □

We have defined *between* for both points and rays. The next theorem affirms that those definitions are consistent with each other.

Theorem 3.3.10. *Let A, B, and C be three noncollinear points and let D be a point on the line $\overleftrightarrow{BC}$. The point D is between points B and C if and only if the ray $\overrightarrow{AD}$ is between rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$.*

Proof. Let A, B, and C be three noncollinear points and let D be a point on $\overleftrightarrow{BC}$ (hypothesis). Suppose, first, that D is between B and C (hypothesis). Then C and D are on the same side of $\overleftrightarrow{AB}$ (Theorem 3.3.9) and B and D are on the same side of $\overleftrightarrow{AC}$ (Theorem 3.3.9 again). Therefore, D is in the interior of $\angle BAC$ (definition of interior of angle) and $\overrightarrow{AD}$ is between $\overrightarrow{AB}$ and $\overrightarrow{AC}$ (definition of betweenness for rays). This completes the proof of the first half of the theorem.

Now suppose the ray $\overrightarrow{AD}$ is between the rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ (hypothesis). Then D is in the interior of $\angle BAC$ (betweenness of rays is well defined). Therefore, B and D are on the same side of $\overleftrightarrow{AC}$ (definition of interior of angle), so C is not on segment $\overline{BD}$ (Plane Separation Postulate). In a similar way we know that B is not on segment $\overline{CD}$. Thus B, C, and D are three collinear points such that B is not between C and D and C is not between B and D. It follows that D is between B and C (Corollary 3.2.19). This completes the proof of the second half of the theorem. □

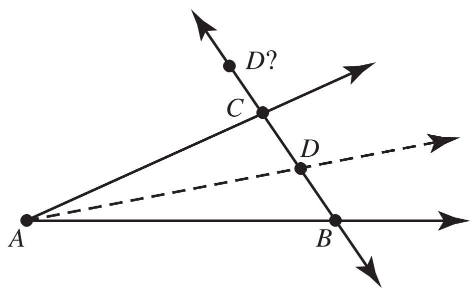

FIGURE 3.15: The point D is between B and C if and only if D is in the interior of $\angle BAC$

The Plane Separation Postulate can be reformulated as a statement about how lines and triangles intersect. First we must provide a definition of triangle.

Definition 3.3.11. Let A, B, and C be three noncollinear points. The *triangle* $\triangle ABC$ consists of the union of the three segments $\overline{AB}$, $\overline{BC}$, and $\overline{AC}$; that is,

$$\triangle ABC = \overline{AB} \cup \overline{BC} \cup \overline{AC}.$$

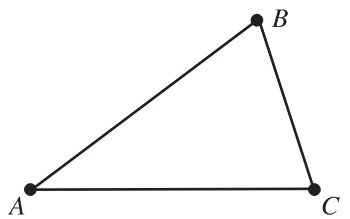

FIGURE 3.16: A triangle

The points A, B, and C are called the *vertices* of the triangle and the segments $\overline{AB}$, $\overline{BC}$, and $\overline{AC}$ are called the *sides* of the triangle.

Notice that we have only defined $\triangle ABC$ in case A, B, and C are noncollinear points. Thus the statement "$\triangle ABC$ is a triangle" is to be interpreted to mean, among other things, that the points A, B, and C are noncollinear.

The following theorem is another example of a result that is labeled an axiom even though we prove it as a theorem. It is called "Pasch's Axiom" because in his formulation of the foundations of geometry, Moritz Pasch (1843–1930) introduced it as an axiom. It is not one of the axioms in our axiomatic systemization of geometry; instead we prove it as a consequence of the Plane Separation Postulate. In fact the proof reveals that it is really just a restatement of parts of the Plane Separation Postulate in different terminology.

Theorem 3.3.12 (Pasch's Axiom). *Let $\triangle ABC$ be a triangle and let ℓ be a line such that none of A, B, and C lies on ℓ. If ℓ intersects $\overline{AB}$, then ℓ also intersects either $\overline{AC}$ or $\overline{BC}$.*

Proof. Let $\triangle ABC$ be a triangle and ℓ be a line such that ℓ intersects $\overline{AB}$ and none of the points A, B, and C lies on ℓ (hypothesis). Let H_1 and H_2 be the two half-planes determined by ℓ (Axiom 3.3.2). The points A and B are in opposite half-planes (hypothesis and Proposition 3.3.4). Let us say that $A \in H_1$ and $B \in H_2$ (notation). It must be the case that either C is in H_1 or C is in H_2 (Axiom 3.3.2). If $C \in H_2$, then $\overline{AC} \cap \ell \neq \emptyset$ (Axiom 3.3.2, Part 2). If $C \in H_1$, then $\overline{BC} \cap \ell \neq \emptyset$ (Axiom 3.3.2, Part 2). □

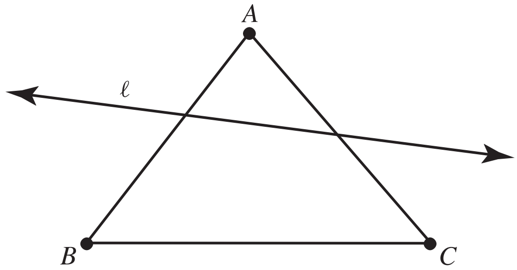

FIGURE 3.17: Pasch's Axiom

EXERCISES 3.3

1. Prove that the intersection of two convex sets is convex. Show by example that the union of two convex sets need not be convex. Is the empty set convex?
2. Let A and B be two distinct points. Prove that each of the sets $\{A\}$, $\overline{AB}$, $\overrightarrow{AB}$, and $\overleftrightarrow{AB}$ is a convex set.
3. Let ℓ be a line and let H be one of the half-planes bounded by ℓ. Prove that $H \cup \ell$ is a convex set.
4. Betweenness on the sphere. In Exercise 3.2.6, betweenness and segments were defined for the sphere $\mathbb{S}^2$. Define *ray* in the usual way (Definition 3.2). A great circle divides the sphere into two hemispheres, which we could think of as half-planes determined by the great circle.
   (a) Does the Plane Separation Postulate hold in this setting? Explain.
   (b) Let A and B be distinct, nonantipodal points on $\mathbb{S}^2$. Find all points C such that $A * B * C$. Sketch the ray $\overrightarrow{AB}$.
   (c) Does Theorem 3.3.9 hold in this setting? Explain.
5. Suppose $\triangle ABC$ is a triangle and ℓ is a line such that none of the vertices A, B, or C lies on ℓ. Prove that ℓ cannot intersect all three sides of $\triangle ABC$. Is it possible for a line to intersect all three sides of a triangle?
6. Prove that the vertex of an angle is well defined (i.e., if $\angle BAC = \angle EDF$, then $A = D$).
7. Prove that the vertices and edges of a triangle are well defined (i.e., if $\triangle ABC = \triangle DEF$, then $\{A, B, C\} = \{D, E, F\}$).

3.4 ANGLE MEASURE AND THE PROTRACTOR POSTULATE

The fifth axiom spells out the properties of angle measure (the last undefined term).

Axiom 3.4.1 (The Protractor Postulate). *For every angle $\angle BAC$ there is a real number $\mu(\angle BAC)$, called the measure of $\angle BAC$, such that the following conditions are satisfied.*

1. $0^\circ \leqslant \mu(\angle BAC) < 180^\circ$ *for every angle* $\angle BAC$.

2. $\mu(\angle BAC) = 0°$ *if and only if* $\overrightarrow{AB} = \overrightarrow{AC}$.
3. (Angle Construction Postulate) *For each real number r,* $0 < r < 180$, *and for each half-plane H bounded by* $\overleftrightarrow{AB}$ *there exists a unique ray* $\overrightarrow{AE}$ *such that E is in H and* $\mu(\angle BAE) = r°$.
4. (Angle Addition Postulate) *If the ray* $\overrightarrow{AD}$ *is between rays* $\overrightarrow{AB}$ *and* $\overrightarrow{AC}$, *then*

$$\mu(\angle BAD) + \mu(\angle DAC) = \mu(\angle BAC).$$

In Part 3 of the postulate, it is the *ray* that is unique, not the point E. Many different points determine the same ray. See Figure 3.13 for an illustration of the angles in Part 4.

We are familiar with the measurement of angles from previous mathematics courses and it is intuitively clear that a measurement function like that described in the Protractor Postulate exists in the Cartesian plane. It is not easy to write down a precise formula, however, and we make no attempt to do so at this point. Most of this book is devoted to an axiomatic investigation of geometry; so it is not necessary to have a definition of angle measure, we simply assume that angle measure exists and work from there. In Chapter 11 we will come back to the question of the existence of models and at that time we will define angle measure in the Cartesian plane.

Even though we do not give a definition of angle measure, we can draw diagrams that illustrate our intuitive understanding. For example, in Figure 3.18 we have $\mu(\angle BAC) = 20°$, $\mu(\angle BAD) = 45°$, $\mu(\angle BAE) = 140°$, and (by the Angle Addition Postulate) $\mu(\angle CAD) = 25°$.

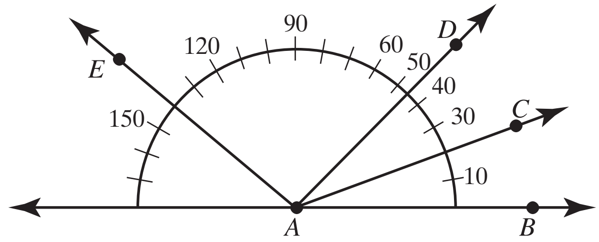

FIGURE 3.18: Measuring angles

The units of angle measure that we use are called *degrees* and are denoted by the symbol °. Just as it is possible to measure distances using various different metrics, so there are other choices of angle measurement function. For example, radian measure is used in many mathematics courses and you are familiar with it from your study of calculus. It is traditional to use degree measure in a geometry course such as this. Degrees are dimensionless and the measure of an angle is really just a real number; the symbol ° merely indicates which angle measurement function is being used. Including the degree symbol occasionally becomes awkward and we will sometimes find it convenient to omit the symbol ° and just use the number.

Congruence of angles is defined using angle measure in much the same way that congruence of segments was defined using distance.

Definition 3.4.2. Two angles $\angle BAC$ and $\angle EDF$ are said to be *congruent*, written $\angle BAC \cong \angle EDF$, if $\mu(\angle BAC) = \mu(\angle EDF)$.

The Protractor Postulate tells us that every angle has a measure that is a real number less than 180°. We will divide angles into three types, depending on how the measure compares with 90°.

Definition 3.4.3. Angle $\angle BAC$ is a *right angle* if $\mu(\angle BAC) = 90°$, $\angle BAC$ is an *acute angle* if $\mu(\angle BAC) < 90°$, and $\angle BAC$ is an *obtuse angle* if $\mu(\angle BAC) > 90°$.

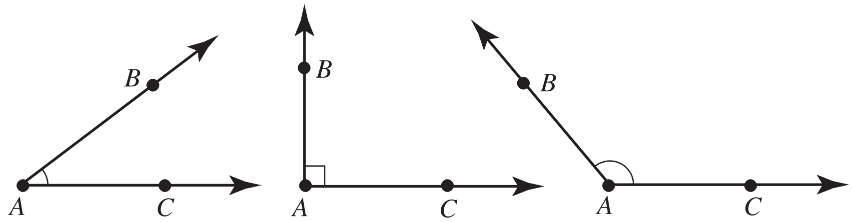

FIGURE 3.19: Types of angles: acute, right, and obtuse

It is worth noting that there are many parallels between the Ruler Postulate and the Protractor Postulate. Both postulates assert that certain geometric properties can be quantified and measured. In fact the postulates are named after the everyday tools we use to measure those quantities: the ruler in case of distance and the protractor in case of angles. Euclid's tools, the compass and the straightedge, have been replaced by new ones. The new tools are like the old ones in that one of them is used on lines and the other on angles (or circles), but they are very unlike the old ones in that both include numerical scales.

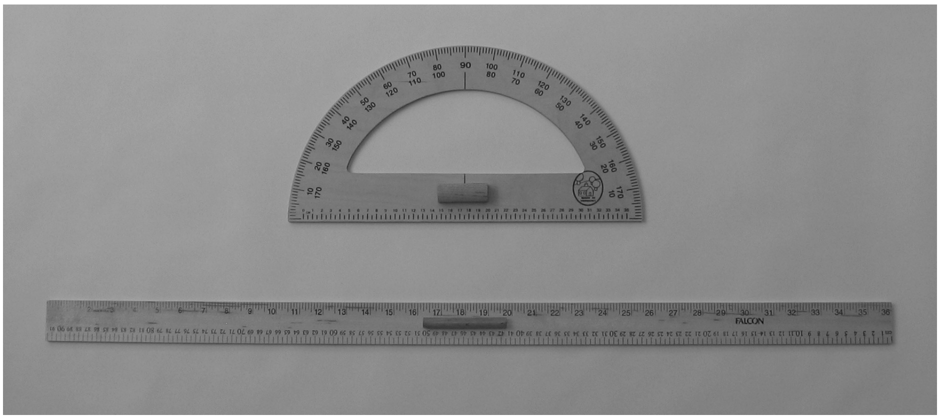

FIGURE 3.20: A protractor and a ruler

The Protractor Postulate can be used to establish a one-to-one correspondence between angles based on one side of the ray $\overrightarrow{AB}$ and the real numbers in the interval $(0, 180)$— see Exercise 2. In fact, G. D. Birkhoff, who was the first to introduce a version of the Protractor Postulate, stated it in terms of just such a one-to-one correspondence (see Appendix B). We choose to state the postulate the way we do because we want to stay as close as possible to the statements of the postulates in high school texts such as [40] and [45].

Betweenness for points was defined in terms of distance, but betweenness for rays was defined using only plane separation. We conclude this section with a theorem that assures us that betweenness of rays is related to angle measure in the expected way. Be sure to note the close analogy between Theorem 3.4.5 and Corollary 3.2.18. The proof of the theorem is clearer if the following technical issue is dealt with separately.

Lemma 3.4.4. *If A, B, C, and D are four distinct points such that C and D are on the same side of* $\overleftrightarrow{AB}$ *and D is not on* $\overrightarrow{AC}$*, then either C is in the interior of* $\angle BAD$ *or D is in the interior of* $\angle BAC$*.*

Proof. Let A, B, C, and D be four distinct points such that C and D are on the same side of $\overleftrightarrow{AB}$ and D is not on $\overrightarrow{AC}$ (hypothesis). We will assume that D is not in the interior of $\angle BAC$ (hypothesis) and use that assumption to prove that C is in the interior of $\angle BAD$.

We know that C and D are on the same side of $\overleftrightarrow{AB}$ and D is not in the interior of $\angle BAC$ (hypothesis), so B and D must lie on opposite sides of $\overleftrightarrow{AC}$ (negation of definition of angle interior). Thus $\overline{BD} \cap \overleftrightarrow{AC} \neq \emptyset$ (Plane Separation Postulate). Let C' be the unique point at which $\overline{BD}$ intersects $\overleftrightarrow{AC}$. (See Figure 3.21.) By Theorem 3.3.10, C' is in the interior of $\angle BAD$. In particular, D and C' lie on the same side of $\overleftrightarrow{AB}$. Since C and C' both lie on the same side of $\overleftrightarrow{AB}$ as D, A cannot be between C and C', so $\overrightarrow{AC}$ and $\overrightarrow{AC'}$ cannot be opposite rays. Therefore, $\overrightarrow{AC} = \overrightarrow{AC'}$. It follows that C is in the interior of $\angle BAD$ (Theorem 3.3.9). □

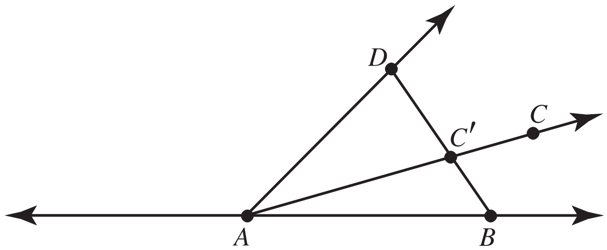

FIGURE 3.21: If D is not in the interior of $\angle BAC$, then C is in the interior of $\angle BAD$

Theorem 3.4.5 (Betweenness Theorem for Rays). *Let A, B, C, and D be four distinct points such that C and D lie on the same side of* $\overleftrightarrow{AB}$*. Then* $\mu(\angle BAD) < \mu(\angle BAC)$ *if and only if* $\overrightarrow{AD}$ *is between rays* $\overrightarrow{AB}$ *and* $\overrightarrow{AC}$*.*

Proof. Let A, B, C, and D be four distinct points such that C and D lie on the same side of $\overleftrightarrow{AB}$ (hypothesis). Assume, first, that $\overrightarrow{AD}$ is between rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ (hypothesis). Then $\mu(\angle BAD) + \mu(\angle DAC) = \mu(\angle BAC)$ (Protractor Postulate, Part 4) and $\mu(\angle DAC) > 0$ (Protractor Postulate, Parts 1 and 2), so $\mu(\angle BAD) < \mu(\angle BAC)$ (algebra).

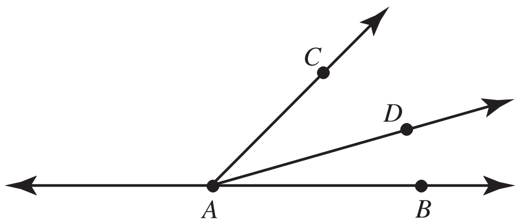

FIGURE 3.22: $\mu(\angle BAD) < \mu(\angle BAC)$ if and only if D is in the interior of $\angle BAC$

We will prove the contrapositive of the second half of the theorem. Suppose $\overrightarrow{AD}$ is not between rays $\overrightarrow{AB}$ and $\overrightarrow{AC}$ (hypothesis). We must prove that $\mu(\angle BAD) \geqslant \mu(\angle BAC)$.

If D lies on $\overrightarrow{AC}$, then $\mu(\angle BAD) = \mu(\angle BAC)$. Otherwise, C is in the interior of $\angle BAD$ (Lemma 3.4.4). Therefore, $\mu(\angle BAD) > \mu(\angle BAC)$ (by the first half of the theorem). □

The Betweenness Theorem for Rays is used in the proof of the existence of angle bisectors just as the Betweenness Theorem of Points was used in the proof of the existence of midpoints.

Definition 3.4.6. Let A, B, and C be three noncollinear points. A ray $\overrightarrow{AD}$ is an *angle bisector* of $\angle BAC$ if D is in the interior of $\angle BAC$ and $\mu(\angle BAD) = \mu(\angle DAC)$.

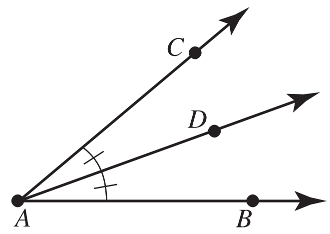

FIGURE 3.23: The angle bisector

Theorem 3.4.7 (Existence and Uniqueness of Angle Bisectors). *If A, B, and C are three noncollinear points, then there exists a unique angle bisector for $\angle BAC$.*

Proof. Exercise 1. □

EXERCISES 3.4

1. Prove existence and uniqueness of angle bisectors (Theorem 3.4.7).

2. Let A and B be two points, let H be one of the half-planes of $\ell = \overleftrightarrow{AB}$, and let $\mathscr{A} = \{\angle BAE \mid E \in H\}$. Define $f : \mathscr{A} \to (0, 180)$ by $f(\angle BAE) = \mu(\angle BAE)$.
   (a) Prove that f is a one-to-one correspondence.
   (b) Prove that $\overrightarrow{AF}$ is between $\overrightarrow{AB}$ and $\overrightarrow{AE}$ if and only if $f(\angle BAF)$ is between 0 and $f(\angle BAE)$.

3.5 THE CROSSBAR THEOREM AND THE LINEAR PAIR THEOREM

Before proceeding to the final postulate, we pause to prove two fundamental theorems of plane geometry, the Crossbar Theorem and the Linear Pair Theorem. Both are often taken as axioms. Readers who are anxious to move more quickly past the foundations can also do that: it is logically acceptable to take the statements as additional axioms and to move on. The proofs can be revisited later when the reader is in a better position to appreciate the power and importance of the results. The main goal we wish to accomplish in this chapter is to lay out explicitly all the basic facts that Euclid took for granted in his proofs; the goal of making those assumptions explicit is much more important that the secondary goal of trying to find a minimal set of necessary assumptions.

The proof of the first theorem relies heavily on the following preliminary result, which is known as the "Z-Theorem" because of the shape of the diagram that accompanies it. The Z-Theorem is an easy consequence of Theorem 3.3.9.

Theorem 3.5.1 (The Z-Theorem). *Let ℓ be a line and let A and D be distinct points on ℓ. If B and E are points on opposite sides of ℓ, then $\overrightarrow{AB} \cap \overrightarrow{DE} = \emptyset$.*

Proof. Except for endpoints, all the points of $\overrightarrow{AB}$ lie in one half-plane and all the points of $\overrightarrow{DE}$ lie in the other half-plane (Theorem 3.3.9). The half-planes are disjoint by the Plane Separation Postulate. Thus the only place the rays could intersect is in their endpoints. But the endpoints are distinct by hypothesis. □

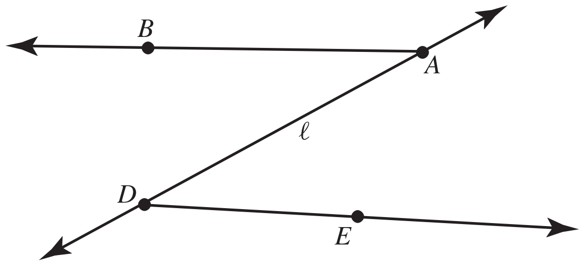

FIGURE 3.24: The rays are disjoint

The Crossbar Theorem asserts that if a ray is in the interior of one of the angles of a triangle, then it must intersect the opposite side of the triangle. The opposite side forms a "crossbar" for the angle (see Figure 3.25).

Theorem 3.5.2 (The Crossbar Theorem). *If $\triangle ABC$ is a triangle and D is a point in the interior of $\angle BAC$, then there is a point G such that G lies on both $\overrightarrow{AD}$ and $\overline{BC}$.*

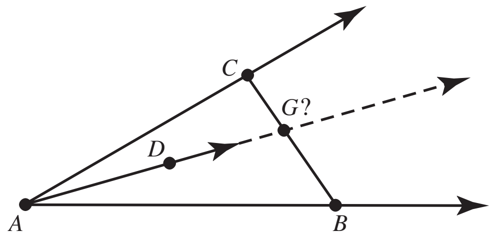

FIGURE 3.25: Statement of Crossbar Theorem

Proof. Let $\triangle ABC$ be a triangle and let D be a point in the interior of $\angle BAC$ (hypothesis). Choose points E and F such that $E * A * B$ and $F * A * D$ (Ruler Postulate) and let $\ell = \overleftrightarrow{AD}$.

Since D is in the interior of $\angle BAC$, neither B nor C lies on ℓ. Thus we can apply Pasch's Axiom (Theorem 3.3.12) to the triangle $\triangle EBC$ to conclude that ℓ must intersect either $\overline{EC}$ or $\overline{BC}$. In order to complete the proof, we must show that it is the ray $\overrightarrow{AD}$ that intersects either $\overline{EC}$ or $\overline{BC}$ (and not the opposite ray $\overrightarrow{AF}$) and that $\overrightarrow{AD}$ does not intersect $\overline{EC}$. In symbols, we must show $\overrightarrow{AF} \cap \overline{EC} = \emptyset$, $\overrightarrow{AF} \cap \overline{BC} = \emptyset$, and $\overrightarrow{AD} \cap \overline{EC} = \emptyset$. This will be accomplished by three applications of the Z-Theorem (Corollary 3.5.1).

Because A is between F and D, F and D lie on opposite sides of $\overleftrightarrow{AB}$ (Plane Separation Postulate). On the other hand, C and D are on the same side of $\overleftrightarrow{AB}$ (D is in the interior of $\angle BAC$), so C and F are on opposite sides of $\overleftrightarrow{AB}$ (Plane Separation

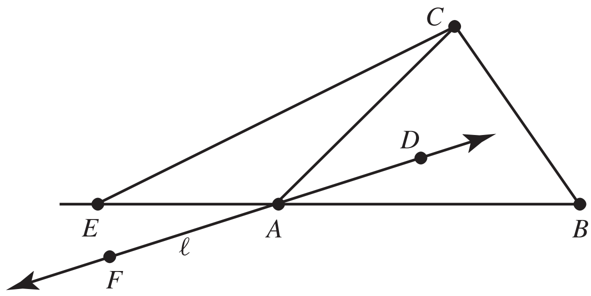

FIGURE 3.26: Proof of the Crossbar Theorem

Postulate). Thus $\overrightarrow{EC} \cap \overrightarrow{AF} = \emptyset$ (Z-Theorem). Since $\overline{EC}$ is a subset of $\overrightarrow{EC}$, we have $\overrightarrow{AF} \cap \overline{EC} = \emptyset$.

By the previous paragraph, C and F are on opposite sides of $\overleftrightarrow{AB}$. Thus $\overrightarrow{BC} \cap \overrightarrow{AF} = \emptyset$ (the Z-Theorem again). Since $\overline{BC}$ is a subset of $\overrightarrow{BC}$, we have $\overrightarrow{AF} \cap \overline{BC} = \emptyset$.

Now A is between E and B, so E and B lie on opposite sides of $\overleftrightarrow{AC}$ (Plane Separation Postulate). On the other hand, B and D are on the same side of $\overleftrightarrow{AC}$ (D is in the interior of $\angle BAC$), so E and D are on opposite sides of $\overleftrightarrow{AC}$ (Plane Separation Postulate). Thus $\overrightarrow{CE} \cap \overrightarrow{AD} = \emptyset$ (the Z-Theorem once again). Since $\overline{EC}$ is a subset of $\overrightarrow{CE}$, we have $\overrightarrow{AD} \cap \overline{EC} = \emptyset$. This completes the proof of the Crossbar Theorem. □

Theorem 3.3.10 asserts that if the ray $\overrightarrow{AD}$ intersects the interior of the crossbar $\overline{BC}$, then D is in the interior of $\angle BAC$. Thus there is a sense in which Theorem 3.3.10 is a converse to the Crossbar Theorem. The next theorem combines them into one theorem.

Theorem 3.5.3. *A point D is in the interior of the angle $\angle BAC$ if and only if the ray $\overrightarrow{AD}$ intersects the interior of the segment $\overline{BC}$.*

Proof. Suppose, first, that D is in the interior of $\angle BAC$. Then $\overrightarrow{AD}$ intersects the interior of $\overline{BC}$ by the Crossbar Theorem. Next suppose that $\overrightarrow{AD}$ intersects the interior of $\overline{BC}$. Let E be a point of intersection. Then E is in the interior of $\angle BAC$ by Theorem 3.3.10 and so D is in the interior of $\angle BAC$ by Theorem 3.3.9. □

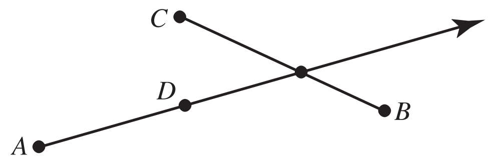

FIGURE 3.27: D is in the interior of $\angle BAC$

Theorem 3.5.3 is taken as an axiom by MacLane [32]. He calls it the "Continuity Axiom" because it implies the Continuity Axiom of Birkhoff—see Theorem 3.5.15. In MacLane's formulation of the axioms, Theorem 3.5.3 is assumed in place of the Plane Separation Postulate.

The second major theorem of the section, the Linear Pair Theorem, is a sort of generalization of the Angle Addition Postulate (Part 4 of the Protractor Postulate). It is often taken as an axiom (see [34], [40], [28], and [48], for example). A common name for it is the *Supplement Postulate*. It is another example of a redundant axiom in the SMSG system. First we need a definition.

Definition 3.5.4. Two angles $\angle BAD$ and $\angle DAC$ form a *linear pair* if $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are opposite rays.

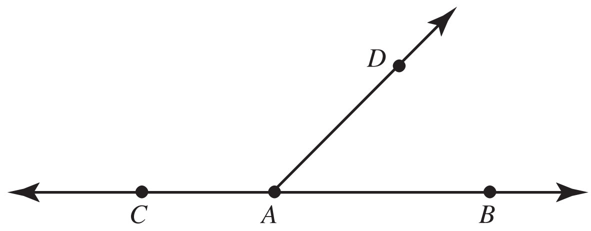

FIGURE 3.28: Angles $\angle BAD$ and $\angle DAC$ form a linear pair

Theorem 3.5.5 (Linear Pair Theorem). *If angles $\angle BAD$ and $\angle DAC$ form a linear pair, then $\mu(\angle BAD) + \mu(\angle DAC) = 180°$.*

Definition 3.5.6. Two angles $\angle BAC$ and $\angle EDF$ are *supplementary* (or *supplements*) if $\mu(\angle BAC) + \mu(\angle EDF) = 180°$.

The Linear Pair Theorem can be restated as follows: *If two angles form a linear pair, then they are supplements.* In order to be a linear pair, two angles must share a side, so supplementary angles do not usually form a linear pair.

Before tackling the full proof of the Linear Pair Theorem, we prove a lemma. The lemma really constitutes part of the proof of the Linear Pair Theorem, but the proof is easier to understand if this technical fact is separated out.

Lemma 3.5.7. *If $C * A * B$ and D is in the interior of $\angle BAE$, then E is in the interior of $\angle DAC$.*

Proof. Let A, B, C, D, and E be five points such that $C * A * B$ and D is in the interior of $\angle BAE$ (hypothesis). In order to show that E is in the interior of $\angle DAC$, we must show that E and D are on the same side of $\overleftrightarrow{AC}$ and that E and C are on the same side of $\overleftrightarrow{AD}$ (definition of angle interior).

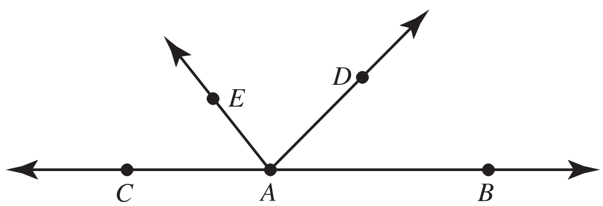

FIGURE 3.29: If D is in the interior of $\angle BAE$, then E is in the interior of $\angle DAC$.

Since D is in the interior of $\angle BAE$, E and D are on the same side of $\overleftrightarrow{AB}$ (definition of angle interior). But $\overleftrightarrow{AB} = \overleftrightarrow{AC}$, so E and D are on the same side of $\overleftrightarrow{AC}$. In addition, $\overrightarrow{AD}$ must intersect $\overline{BE}$ (Crossbar Theorem). Therefore, E and B lie on opposite sides of

$\overleftrightarrow{AD}$ (Plane Separation Postulate). The fact that A is between C and B means that C and B are on opposite sides of $\overleftrightarrow{AD}$ and so we can conclude that C and E are on the same side of $\overleftrightarrow{AD}$ (Plane Separation Postulate). □

Proof of the Linear Pair Theorem. Let $\angle BAD$ and $\angle DAC$ be two angles that form a linear pair (hypothesis). Then $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are opposite rays. In order to simplify notation in the proof, we will drop the degree notation and use α to denote $\mu(\angle BAD)$ and β to denote $\mu(\angle DAC)$. We must prove that $\alpha + \beta = 180$. By trichotomy, either $\alpha + \beta < 180$, $\alpha + \beta = 180$, or $\alpha + \beta > 180$. We will show that the first and last possibilities lead to contradictions, and this will allow us to conclude that $\alpha + \beta = 180$.

Suppose, first, that $\alpha + \beta < 180$. By the Angle Construction Postulate (Protractor Postulate, Part 3), there is a point E, on the same side of $\overleftrightarrow{AB}$ as D, such that $\mu(\angle BAE) = \alpha + \beta$. By the Betweenness Theorem for Rays (Theorem 3.4.5), D is in the interior of $\angle BAE$. Therefore, $\mu(\angle BAD) + \mu(\angle DAE) = \mu(\angle BAE)$ (Angle Addition Postulate) and so $\mu(\angle DAE) = \beta$ (algebra). Now E is in the interior of $\angle DAC$ (Lemma 3.5.7), so $\mu(\angle DAE) + \mu(\angle EAC) = \mu(\angle DAC)$ (Angle Addition Postulate). It follows that $\mu(\angle EAC) = 0$ (algebra). This contradicts Parts 1 and 2 of the Protractor Postulate. Thus we conclude that $\alpha + \beta < 180$ is impossible.

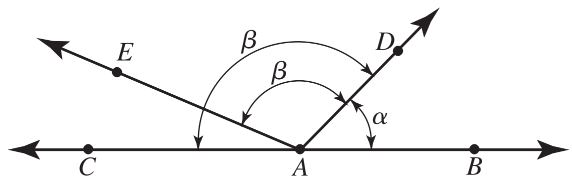

FIGURE 3.30: Proof of the Linear Pair Theorem in case $\alpha + \beta < 180$

Suppose, next, that $\alpha + \beta > 180$. Choose a point F, on the same side of $\overleftrightarrow{AB}$ as D, such that $\mu(\angle BAF) = (\alpha + \beta) - 180$ (Protractor Postulate, Part 3). Now $\beta < 180$ (Protractor Postulate, Part 1), so $\alpha + \beta - 180 < \alpha$ (algebra). By the Betweenness Theorem for Rays (Theorem 3.4.5), F is in the interior of $\angle BAD$. Therefore $\mu(\angle BAF) + \mu(\angle FAD) = \mu(\angle BAD)$ (Angle Addition Postulate) and so $\mu(\angle FAD) = 180 - \beta$ (algebra). By Lemma 3.5.7, D is in the interior of $\angle FAC$. Thus $\mu(\angle FAD) + \mu(\angle DAC) = \mu(\angle FAC)$ (Angle Addition Postulate). It follows that $\mu(\angle FAC) = 180^\circ$ (algebra). This contradicts Part 1 of the Protractor Postulate and so we conclude that $\alpha + \beta > 180$ is impossible. This completes the proof of the Linear Pair Theorem. □

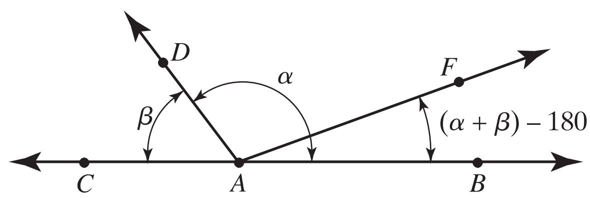

FIGURE 3.31: Proof of the Linear Pair Theorem in case $\alpha + \beta > 180$

Definition 3.5.8. Two lines ℓ and m are *perpendicular* if there exists a point A that lies on both ℓ and m and there exist points $B \in \ell$ and $C \in m$ such that $\angle BAC$ is a right angle. Notation: $\ell \perp m$.

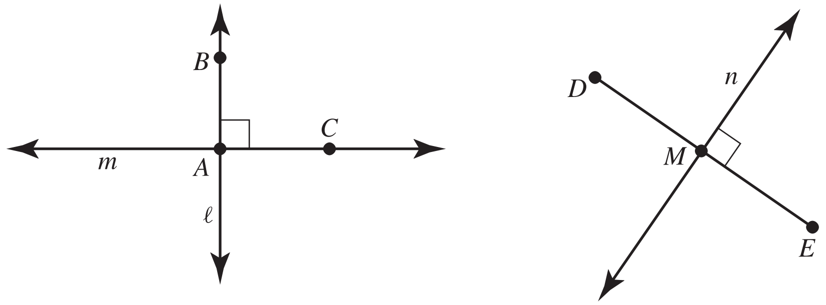

FIGURE 3.32: ℓ and m are perpendicular lines; n is the perpendicular bisector of $\overline{DE}$

Theorem 3.5.9. *If ℓ is a line and P is a point on ℓ, then there exists exactly one line m such that P lies on m and $m \perp \ell$.*

Proof. Exercise 2. □

Definition 3.5.10. Let D and E be two distinct points. A *perpendicular bisector* of $\overline{DE}$ is a line n such that the midpoint of $\overline{DE}$ lies on n and $n \perp \overleftrightarrow{DE}$.

Theorem 3.5.11 (Existence and Uniqueness of Perpendicular Bisectors). *If D and E are two distinct points, then there exists a unique perpendicular bisector for $\overline{DE}$.*

Proof. Exercise 3. □

Definition 3.5.12. Angles $\angle BAC$ and $\angle DAE$ form a *vertical pair* (or are *vertical angles*) if rays $\overrightarrow{AB}$ and $\overrightarrow{AE}$ are opposite and rays $\overrightarrow{AC}$ and $\overrightarrow{AD}$ are opposite or if rays $\overrightarrow{AB}$ and $\overrightarrow{AD}$ are opposite and rays $\overrightarrow{AC}$ and $\overrightarrow{AE}$ are opposite.

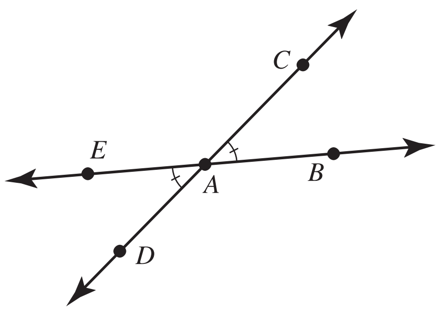

FIGURE 3.33: $\angle BAC$ and $\angle DAE$ are vertical angles

Theorem 3.5.13 (Vertical Angles Theorem). *Vertical angles are congruent.*

Proof. Exercise 5. □

We conclude this section with an application of the Crossbar Theorem. The final theorem in this section will not be used again until we prove circular continuity in Chapter 8, so the remainder of the section can be omitted for now without serious consequence.

The theorem we will prove is called the *Continuity Axiom*. It asserts that the relationship between angle measure and distance is a continuous one. It is called an axiom because Birkhoff [4] stated this fact as part of his version of the Protractor Postulate. It is included here because it adds interest to the statement of the Crossbar Theorem and because it relates the Crossbar Theorem to the more familiar concept of continuity.

We first need a lemma about functions defined on intervals of real numbers. The lemma should seem intuitively plausible: If a function from the real numbers to the real numbers is strictly increasing, then a jump in the graph would result in a gap in the range of the function.

Lemma 3.5.14. *Let $[a, b]$ and $[c, d]$ be closed intervals of real numbers and let $f : [a, b] \to [c, d]$ be a function. If f is strictly increasing and onto, then f is continuous.*

Proof. Let $f : [a, b] \to [c, d]$ be a function that is both increasing and onto (hypothesis) and let $x \in [a, b]$. We must show that f is continuous at x. Let us assume that x is in the interior of $[a, b]$. The proof of the case in which x is an endpoint is similar.

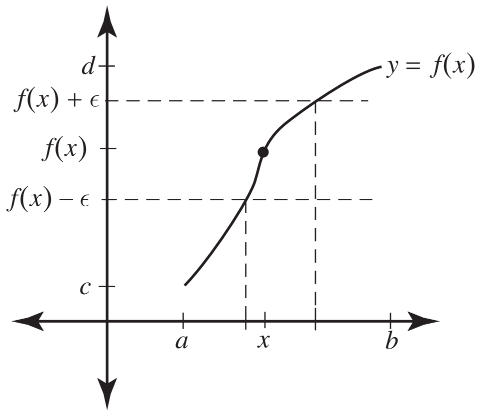

FIGURE 3.34: Proof of Lemma 3.5.14

Let $\epsilon > 0$ be given. We must show that there exists a positive number δ such that if $|x - y| < \delta$, then

$$|f(x) - f(y)| < \epsilon$$

(definition of continuous at x).

Since f is strictly increasing, $c < f(x) < d$. We may assume that $c < f(x) - \epsilon$ and $f(x) + \epsilon < d$. (If this is not the case, simply replace ϵ with a smaller number.) Since f is onto, there exist numbers x_1 and x_2 in $[a, b]$ such that $f(x_1) = f(x) - \epsilon$ and $f(x_2) = f(x) + \epsilon$. Choose $\delta > 0$ to be small enough so that $x_1 < x - \delta$ and $x_2 > x + \delta$.

Suppose $|x - y| < \delta$. Then $x - \delta < y < x + \delta$. Since f is increasing, $f(x_1) < f(x - \delta) < f(y) < f(x + \delta) < f(x_2)$. Therefore, $f(x) - \epsilon < f(y) < f(x) + \epsilon$ and so $|f(x) - f(y)| < \epsilon$ (algebra). □

Setting for the Continuity Axiom. Let A, B, and C be three noncollinear points. For each point D on $\overline{BC}$ there is an angle $\angle CAD$ and there is a distance CD. We will

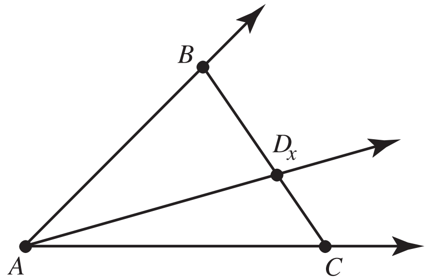

FIGURE 3.35: The Continuity Axiom

define a function that relates the distance and the angle measure. Let $d = BC$. The Ruler Placement Postulate (Theorem 3.2.16) gives a one-to-one correspondence from the interval $[0, d]$ to points on $\overline{BC}$ such that C corresponds to 0 and B corresponds to d. Let D_x be the point that corresponds to the number x; i.e., D_x is the point on $\overline{BC}$ such that $CD_x = x$. Define a function $f : [0, d] \to [0, \mu(\angle CAB)]$ by $f(x) = \mu(\angle CAD_x)$.

Theorem 3.5.15 (The Continuity Axiom). *The function f described in the previous paragraph is a continuous function, as is the inverse of f.*

Proof. Let f be the function described above. By Theorems 3.3.10 and 3.4.5, f is a strictly increasing function. By the Crossbar Theorem and Theorem 3.4.5, f is onto. Therefore, f is continuous (Lemma 3.5.14). It is obvious that the inverse of f is increasing and onto, so the inverse is also continuous. □

EXERCISES 3.5

1. If $\ell \perp m$, then ℓ and m contain rays that make four different right angles.
2. Prove existence and uniqueness of a perpendicular to a line at a point on the line (Theorem 3.5.9).
3. Prove existence and uniqueness of perpendicular bisectors (Theorem 3.5.11).
4. Prove that supplements of congruent angles are congruent.
5. Restate the Vertical Angles Theorem (Theorem 3.5.13) in if-then form. Prove the theorem.
6. Prove the following converse to the Vertical Angles Theorem: *If A, B, C, D, and E are points such that $A * B * C$, D and E are on opposite sides of $\overleftrightarrow{AB}$, and $\angle DBC \cong \angle ABE$, then D, B, and E are collinear.*
7. Use the Continuity Axiom and the Intermediate Value Theorem to prove the Crossbar Theorem.

3.6 THE SIDE-ANGLE-SIDE POSTULATE

So far we have formulated one axiom for each of the undefined terms. It would be reasonable to expect this to be enough axioms since we now know the basic properties of each of the undefined terms. But there is still something missing: The postulates stated so far do not tell us quite enough about how distance (or length of segments) and angle measure interact with each other. In this section we will give an example that illustrates the need for additional information and then state one final axiom to complete the picture.

The simplest objects that combine both segments and angles are triangles. We have defined what it means for two segments to be congruent and what it means for two angles to be congruent. We now extend that definition to triangles, where the two types of congruence are combined.

Definition 3.6.1. Two triangles are *congruent* if there is a correspondence between the vertices of the first triangle and the vertices of the second triangle such that corresponding angles are congruent and corresponding sides are congruent.

The assertion that two triangles are congruent is really the assertion that there are six congruences, three angle congruences and three segment congruences.

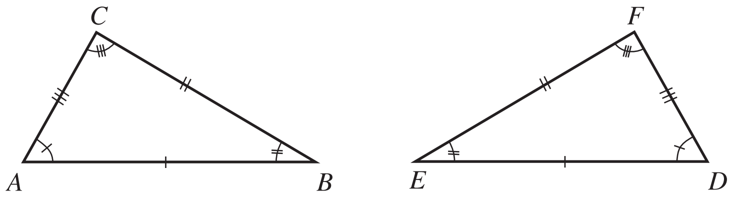

FIGURE 3.36: Congruent triangles

Notation. We use the symbol $\cong$ to indicate congruence of triangles. It is understood that the notation $\triangle ABC \cong \triangle DEF$ means that the two triangles are congruent under the correspondence $A \leftrightarrow D$, $B \leftrightarrow E$, and $C \leftrightarrow F$. Specifically, $\triangle ABC \cong \triangle DEF$ means $\overline{AB} \cong \overline{DE}$, $\overline{BC} \cong \overline{EF}$, $\overline{AC} \cong \overline{DF}$, $\angle ABC \cong \angle DEF$, $\angle BCA \cong \angle EFD$, and $\angle CAB \cong \angle FDE$.

One of the most basic theorems in geometry is the Side-Angle-Side Triangle Congruence Condition (abbreviated SAS). It states that if two sides and the included angle of one triangle are congruent to two sides and included angle of a second triangle, then the triangles are congruent. This theorem is Euclid's Proposition 4 and Euclid's proof of the proposition was discussed in Chapter 1. SAS is one of the essential ingredients in the proofs of more complicated theorems of geometry, so it is a key theorem we need in our development of the subject. But the axioms we have stated so far are not strong enough to allow us to prove SAS as a theorem.

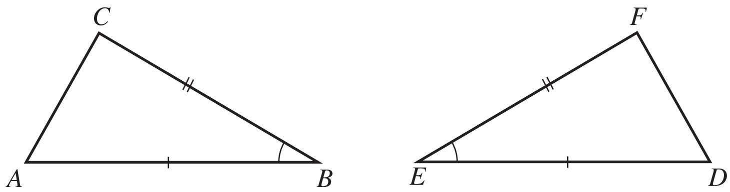

FIGURE 3.37: Side-Angle-Side assures us that these two triangles are congruent

We have seen that the Cartesian plane with the taxicab metric satisfies the Ruler Postulate. While distances are measured in an unusual way in taxicab geometry, the lines are the standard Euclidean lines and so we can use the ordinary Euclidean angle measure to measure angles in taxicab geometry. The Cartesian plane equipped with the taxicab metric and Euclidean angle measure is a model for geometry that satisfies all five of the axioms stated so far in this chapter. The next example shows that SAS fails in that model.

EXAMPLE 3.6.2 SAS fails in taxicab geometry

Consider the triangles $\triangle ABC$ and $\triangle DEF$ in the Cartesian plane that have vertices at $A = (0,0)$, $B = (0,2)$, $C = (2,0)$, $D = (0,1)$, $E = (-1,0)$, and $F = (1,0)$. In the taxicab metric, $AB = AC = DE = DF = 2$. Furthermore, both the angles $\angle BAC$ and $\angle EDF$ are right angles. Nonetheless, $\triangle ABC \ncong \triangle DEF$ because $BC = 4$ while $EF = 2$. ■

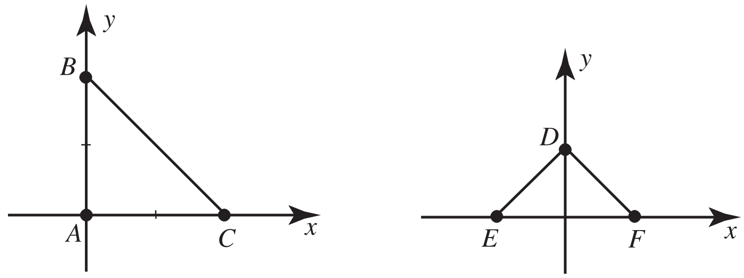

FIGURE 3.38: Two triangles in taxicab geometry

From this example we see that the postulates we have stated thus far are not powerful enough to prove the SAS theorem. We will fix this by simply assuming Side-Angle-Side as an axiom. In doing so we are following both Hilbert [25] and SMSG [40]. Since SAS is true in the Cartesian plane with the Euclidean metric but fails in the Cartesian plane with the taxicab metric, the SAS Postulate is independent of the other postulates of neutral geometry.

Axiom 3.6.3 (The Side-Angle-Side Postulate or SAS). *If $\triangle ABC$ and $\triangle DEF$ are two triangles such that $\overline{AB} \cong \overline{DE}$, $\angle ABC \cong \angle DEF$, and $\overline{BC} \cong \overline{EF}$, then $\triangle ABC \cong \triangle DEF$.*

Euclid proved SAS using his method of superposition, so Euclid's "proof" depends on the ability to move triangles around in the plane without distorting them. Examples such as the one above show that we cannot take this for granted. Even though taxicab geometry satisfies all the axioms stated earlier in this chapter, it allows only a limited number of rigid motions (motions that preserve both distances and angle measures).

Some high school textbooks (such as [45]) replace the SAS Postulate with an axiom which states that certain kinds of transformations are allowed and then prove SAS as a theorem. We will discuss that transformational approach in Chapter 10. It is interesting to note that the latter approach is in some ways closer to Euclid's original treatment than is the standard college-level approach of taking SAS as an axiom. Euclid gives Common Notion 4 ("Things which coincide with one another are equal to one another") as the justification for the method of superposition that he uses to prove SAS. It is not unreasonable to interpret Common Notion 4 to be an axiom that permits rigid motions in the plane.

The transformational approach has implications for the definition of congruence as well. This will also be discussed in more detail in Chapter 10. For now we will simply point out that we have defined congruence of triangles to mean that corresponding parts are congruent. (A triangle has six "parts," three sides and three angles.) In high school textbooks congruence of triangles is defined differently and so the statement "corresponding parts of congruent triangles are congruent" becomes a theorem. That theorem is abbreviated CPCTC, or more generally, CPCFC, which stands for "corresponding parts of congruent figures are congruent." For us, CPCTC is a definition not a theorem.

In order to illustrate the use of SAS, we now give two proofs of the familiar Isosceles Triangle Theorem. The example proofs also serve to illustrate the use of the theorems from the previous section.

Definition 3.6.4. A triangle is called *isosceles* if it has a pair of congruent sides. The two angles not included between the congruent sides are called *base angles*.

Theorem 3.6.5 (Isosceles Triangle Theorem). *The base angles of an isosceles triangle are congruent.*

Restatement. *If $\triangle ABC$ is a triangle and $\overline{AB} \cong \overline{AC}$, then $\angle ABC \cong \angle ACB$.*

First proof of Theorem 3.6.5. Let $\triangle ABC$ be a triangle such that $\overline{AB} \cong \overline{AC}$ (hypothesis). We must prove that $\angle ABC \cong \angle ACB$. Let D be a point in the interior of $\angle BAC$ such that $\overrightarrow{AD}$ is the bisector of $\angle BAC$ (Theorem 3.4.7). There is a point E at which the ray $\overrightarrow{AD}$ intersects the segments $\overline{BC}$ (Crossbar Theorem 3.5.2). Then $\triangle BAE \cong \triangle CAE$ (SAS) and so $\angle ABE \cong \angle ACE$ (definition of congruent triangles). This completes the proof because $\angle ACE = \angle ACB$ and $\angle ABE = \angle ABC$. □

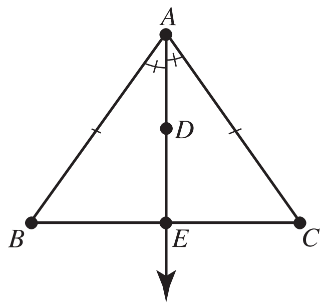

FIGURE 3.39: First proof of the Isosceles Triangle Theorem

The proof above, which is essentially the same as the proof given in a typical high school geometry textbook, has two drawbacks from our point of view. One is the fact that it requires the use of the Crossbar Theorem. Many lower-level textbooks give this proof and simply ignore the fact that the Crossbar Theorem is needed. (The use of the Crossbar Theorem can be suppressed by saying something like this: "Let E be the point at which the angle bisector intersects the side $\overline{BC}$." It is manifestly obvious from the diagram that such a point exists, so most readers do not ask for a justification.) A second drawback to the proof is the fact that it is unnecessarily complicated. There is a much more elegant way to prove the theorem that does not make use of the Crossbar Theorem at all.

Second proof of Theorem 3.6.5. Let $\triangle ABC$ be a triangle such that $\overline{AB} \cong \overline{AC}$ (hypothesis). Then $\triangle BAC \cong \triangle CAB$ (SAS), so $\angle ABC \cong \angle ACB$ (definition of congruent triangles). □

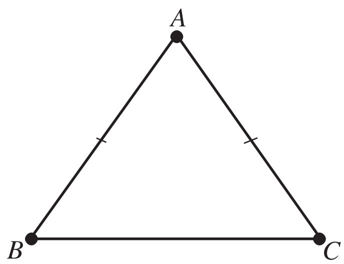

FIGURE 3.40: Second proof of the Isosceles Triangle Theorem

We can understand the second proof visually by imagining that the triangle is picked up, turned over, and placed back on top of itself, bringing A back to where it was but interchanging B and C. This beautifully simple proof is attributed to Pappus of Alexandria who lived about 600 years after Euclid and was the last great ancient Greek geometer. So why do high school textbooks give the longer, more complicated, proof? Mostly because of the transformational approach that is emphasized in those texts. The concept of a reflection across a line is taken to be a more basic concept than the abstract congruence we use in the second proof. The second proof also requires more sophisticated thinking in that we must think of one triangle as two and imagine a correspondence of a triangle to itself that is not the identity. We somehow feel more comfortable thinking of the two triangles in the SAS Postulate as two completely separate triangles, even though there is nothing in the statement that actually requires this.

The Isosceles Triangle Theorem is Euclid's Proposition 5, so we are following in the tradition of Euclid when we give this theorem as the very first application of SAS. Euclid himself gives a proof that is different from either of the two proofs discussed here. Euclid extends the sides $\overline{AB}$ and $\overline{AC}$ first and then works with two triangles that lie below the original triangle. It seems likely that he made this (unnecessary) construction in order to have two distinct triangles to which he could apply SAS.

Euclid's Proposition 5 was known in medieval universities as the *pons asinorum* (the bridge of asses), probably because the diagram that accompanies Euclid's proof looks like a bridge. Another possible explanation for the name is the fact that this proof is a narrow bridge that separates those who can understand and appreciate Euclid's work from those who cannot.

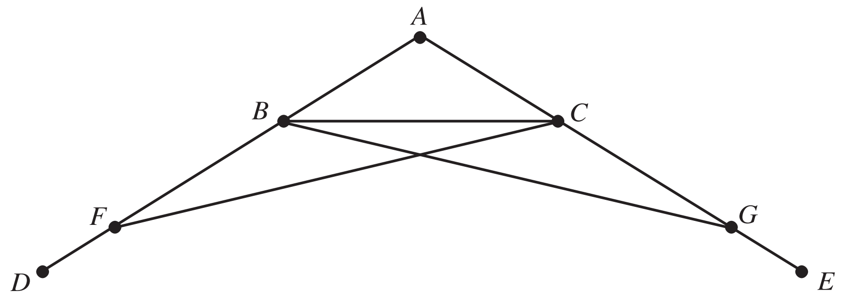

FIGURE 3.41: The *pons asinorum*

EXERCISES 3.6

1. Prove existence and uniqueness of angle bisectors (Theorem 3.4.7) using SAS and the Isosceles Triangle Theorem but not using the Betweenness Theorem for Rays.
2. Recall (Exercise 3.2.8) that the square metric distance between two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ is given by $D((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$. Show by example that $\mathbb{R}^2$ with the square metric and the usual angle measurement function does not satisfy the SAS Postulate.

3.7 THE PARALLEL POSTULATES AND MODELS

The geometry that can be done using only the six postulates stated earlier in this chapter is called *neutral geometry*. As we will see in the next chapter, a great many of the standard theorems of geometry can be proved in neutral geometry. But we will also find that the neutral postulates are not strong enough to allow us to prove much about parallel lines. In Chapter 2 we stated three postulates regarding parallelism. We repeat them here for completeness.

Euclidean Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there is exactly one line m such that P lies on m and $m \parallel \ell$.*

Elliptic Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there is no line m such that P lies on m and $m \parallel \ell$.*

Hyperbolic Parallel Postulate. *For every line ℓ and for every point P that does not lie on ℓ, there are at least two lines m and n such that P lies on both m and n and both m and n are parallel to ℓ.*

We will see in the next chapter that it is possible to prove the existence of parallel lines in neutral geometry. It follows that the Elliptic Parallel Postulate is not consistent with the axioms of neutral geometry. Later in the course we will prove that each of the other two parallel postulates is consistent with the axioms of neutral geometry. We will use the term *Euclidean geometry* to refer to the geometry we can do if we add the Euclidean Parallel Postulate as a seventh postulate and *hyperbolic geometry* to refer to the geometry we can do if we instead add the Hyperbolic Parallel Postulate as a seventh postulate.

The method we will use to prove that the Euclidean and hyperbolic parallel postulates are both consistent with the axioms of neutral geometry is to construct models. We need to develop many more tools before we can construct those models, but we should at least briefly consider the question of the existence of models here.

It is clear from the Ruler Postulate that any model for neutral geometry must contain an infinite number of points, so we can rule out all of the finite geometries of Chapter 2 as possible models for neutral geometry. We can also rule out the sphere $\mathbb{S}^2$ since it does not satisfy either the Incidence Postulate or the Ruler Postulate. The rational plane fails as a model for neutral geometry because it does not satisfy the Ruler Postulate and taxicab geometry fails to satisfy SAS.

This leaves only the Cartesian plane and the Klein disk as possible candidates from among the previously discussed examples. The Cartesian plane with the Euclidean metric certainly is a model for neutral geometry and it is the model you studied in high school. Although it is far from obvious, the Klein disk can also be made into a model for neutral geometry. The undefined terms *point* and *line* have already been interpreted in the Klein disk; in order to see that it is a model for neutral geometry it is necessary to interpret the terms *distance* and *angle measure* and then to verify that all the axioms of neutral geometry are valid in that interpretation. The natural way to measure distances in the Klein disk will certainly not work since there is a finite upper bound on distances and that contradicts the Ruler Postulate. One of the major tasks of the second half of this course is to define a metric in the Klein disk in such a way that the result is a model for hyperbolic geometry.

The question of the existence of models will be set aside for now. First we will give axiomatic treatments of neutral, Euclidean, and hyperbolic geometries. Because we have not verified completely that there is a model for neutral geometry, we also have not proved that the axioms of neutral geometry are consistent. For now we will proceed in the faith that there is a model and will revisit the question of consistency in a later chapter.

Even though we will treat the three geometries axiomatically, we will still want to draw diagrams that illustrate the relationships being discussed. A diagram should be viewed as one possible interpretation of a theorem. Both the theorem statements and the proofs should stand alone and should not depend on the diagrams. The purpose of a diagram is to help us to visually organize the information being discussed. Diagrams add tremendously to our intuitive understanding, but they are only meant to illustrate, not to narrow the range of possible interpretations.

EXERCISES 3.7

1. Check that the trivial geometry containing just one point and no lines satisfies all the postulates of neutral geometry except the Existence Postulate. Which parallel postulate is satisfied by this geometry? (This exercise is supposed to convince you of the need for the Existence Postulate.)

2. Relationship between neutral geometry and incidence geometry.

(a) Explain how the axioms stated in this chapter imply that there exists at least one line.

(b) Explain how the existence of three noncollinear points (Incidence Axiom I-3) can be deduced from the axioms stated in this chapter.

(c) Explain why every model for neutral geometry must have infinitely many points.

(d) Explain why every model for neutral geometry must have infinitely many lines.

(e) Explain why every model for neutral geometry must also be a model for incidence geometry.

(f) Explain why all the theorems of incidence geometry must also be theorems in neutral geometry.

CHAPTER 4

Neutral Geometry

In this chapter we study neutral geometry from an axiomatic point of view (no model is assumed). As explained in the previous chapter, *neutral geometry* is the geometry that is based on the five undefined terms *point*, *line*, *distance*, *half-plane*, and *angle measure* together with the following axioms:

1. The Existence Postulate
2. The Incidence Postulate
3. The Ruler Postulate
4. The Plane Separation Postulate
5. The Protractor Postulate
6. The Side-Angle-Side Postulate

Neutral geometry is ‘‘neutral’’ in the sense that it does not take a stand on the parallel postulate. It is the geometry we can do without any postulate regarding parallelism.

The axioms just named spell out all the assumptions we will use to replace Euclid's first four postulates together with all the unstated assumptions that Euclid took for granted in his proofs. Now we move on to the propositions that Euclid stated and proved in Book I of the *Elements*. We intend to study those propositions while maintaining a ‘‘neutral’’ position regarding parallelism, so our treatment will necessarily differ quite substantially from that of Euclid.

By studying neutral geometry as a separate subject, we are able to clarify the role of the parallel postulate in geometry. The first part of our study of neutral geometry consists of proving as many of the theorems of plane geometry as we can without assuming any axiom regarding parallelism. We will see that a surprisingly large number of the standard theorems from high school geometry can be proved without any parallel postulate at all.

There will be other theorems that cannot be proved without some additional axiom regarding parallelism. In the second part of the chapter we will see that many familiar theorems not only require a parallel postulate for their proofs, but are, in fact, logically equivalent to the Euclidean Parallel Postulate (at least in the context of the axioms of neutral geometry). The chapter ends with a theorem, known as the Universal Hyperbolic Theorem, which asserts that the negation of the Euclidean Parallel Postulate is equivalent in neutral geometry to the Hyperbolic Parallel Postulate. Thus there are only two possibilities: In any model for neutral geometry, either the Euclidean Parallel Postulate holds or the Hyperbolic Parallel Postulate holds.

There is a sense in which Euclid himself did neutral geometry first. He did not use his Fifth Postulate until it was absolutely necessary to do so. It seems clear from his arrangement of the material in Book I of the *Elements* that he recognized that the Fifth Postulate plays a special role in geometry and that it would be good to prove as many theorems as possible without it. In fact, all of the major theorems in the first four sections of this chapter are found in Book I of the *Elements*, and Euclid proves all of them without any mention of his Fifth Postulate.

It is not entirely standard to classify all the material in this chapter as neutral geometry. Many authors include the proofs that various statements are equivalent to the Euclidean Parallel Postulate in their treatment of Euclidean geometry and include the Universal Hyperbolic Theorem in their treatment of hyperbolic geometry. The general organizational principle used in this book is the following: Each theorem has a natural context and that context is an axiomatic system; the axioms of the relevant system are the unstated hypotheses in any theorem. From that perspective it is obvious that all the theorems in this chapter must be classified as theorems in neutral geometry. What they have in common is the fact that the six neutral axioms are implicitly assumed as unstated hypotheses. By way of contrast, when we prove that the Euclidean Parallel Postulate implies some other statement, we will explicitly state, as an additional hypothesis, the fact that the Euclidean Parallel Postulate is being assumed. Similarly, the only unstated hypotheses in the Universal Hyperbolic Theorem are the axioms of neutral geometry.

4.1 THE EXTERIOR ANGLE THEOREM AND EXISTENCE OF PERPENDICULARS

The first major theorem of the chapter, the Exterior Angle Theorem, is one of the fundamental results of neutral geometry. It is Euclid's Proposition 16, and his proof of this proposition is one of the proofs that was studied in Chapter 1. Our "modern" proof is almost exactly the same as Euclid's, right down to the diagram we use to understand the relationships in the proof. There is one important difference however: we use the axioms and theorems of Chapter 3 to fill the gap in Euclid's proof. As you read the proof below, be sure to notice how the postulates and theorems of Chapter 3 provide exactly the information needed to give a complete proof.

Definition 4.1.1. Let $\triangle ABC$ be a triangle. The angles $\angle CAB$, $\angle ABC$, and $\angle BCA$ are called *interior angles* of the triangle. An angle that forms a linear pair with one of the interior angles is called an *exterior angle* for the triangle. If the exterior angle forms a linear pair with the interior angle at one vertex, then the interior angles at the other two vertices are referred to as *remote interior angles*.

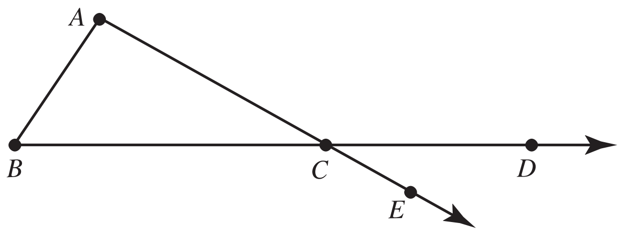

FIGURE 4.1: Angles $\angle ACD$ and $\angle BCE$ are exterior angles for $\triangle ABC$ at vertex C; angles $\angle BAC$ and $\angle ABC$ are remote interior angles

At each vertex of the triangle there are two exterior angles. Those two exterior angles form a vertical pair and are therefore congruent (Theorem 3.5.13).

Theorem 4.1.2 (Exterior Angle Theorem). *The measure of an exterior angle for a triangle is strictly greater than the measure of either remote interior angle.*

Restatement. *If $\triangle ABC$ is a triangle and D is a point such that $\overrightarrow{CD}$ is opposite to $\overrightarrow{CB}$, then $\mu(\angle DCA) > \mu(\angle BAC)$ and $\mu(\angle DCA) > \mu(\angle ABC)$.*

Proof. Let $\triangle ABC$ be a triangle and let D be a point such that $\overrightarrow{CD}$ is opposite to $\overrightarrow{CB}$ (hypothesis). We must prove that $\mu(\angle DCA) > \mu(\angle BAC)$ and that $\mu(\angle DCA) > \mu(\angle ABC)$.

Let E be the midpoint of $\overline{AC}$ (Theorem 3.2.22) and choose F to be the point on $\overrightarrow{BE}$ such that $\overline{BE} \cong \overline{EF}$ (Point Construction Postulate). Notice that $\angle BEA \cong \angle FEC$ (Vertical Angles Theorem). Hence $\triangle BEA \cong \triangle FEC$ (SAS) and so $\angle FCA \cong \angle BAC$ (definition of congruent triangles).

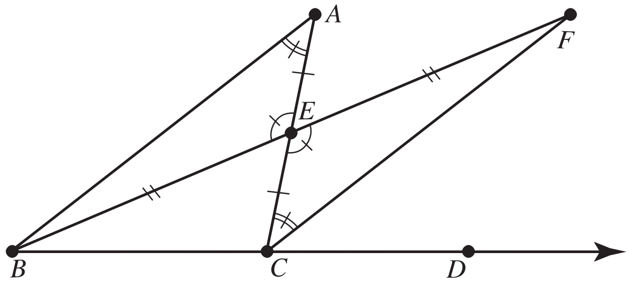

FIGURE 4.2: Proof of the Exterior Angle Theorem

Now F and B are on opposite sides of $\overleftrightarrow{AC}$ and B and D are on opposite sides of $\overleftrightarrow{AC}$, so F and D are on the same side of $\overleftrightarrow{AC}$ (Plane Separation Postulate). Also A and E are on the same side of $\overleftrightarrow{CD}$ (Theorem 3.3.9) and E and F are on the same side of $\overleftrightarrow{CD}$ (Theorem 3.3.9 again), so A and F are on the same side of $\overleftrightarrow{CD}$ (Plane Separation Postulate). The last two sentences show that F is in the interior of $\angle ACD$ (definition of angle interior). Hence $\mu(\angle DCA) > \mu(\angle FCA)$ (Betweenness Theorem for Rays, Theorem 3.4.5). Combining this last statement with the previous paragraph gives $\mu(\angle DCA) > \mu(\angle BAC)$ as required.

To prove $\mu(\angle DCA) > \mu(\angle ABC)$ we proceed in exactly the same way as for the other inequality. The only difference is that we replace the exterior angle $\angle DCA$ with the vertical exterior angle $\angle GCB$, where $\overrightarrow{CG}$ is opposite to $\overrightarrow{CA}$. □

The proof of the Exterior Angle Theorem clearly makes use of the postulates of neutral geometry. This is not just an artifact of the proof we have given, but the theorem itself might fail without those assumptions. Consider, for example, the triangle $\triangle ABC$ on the sphere $\mathbb{S}^2$ that is shown in Figure 4.3. In that triangle, the measure of the interior angle at vertex A can be much larger than 90° while the angles at the other two vertices measure exactly 90°. As a result, the measure of the exterior angle at B or C can be smaller than the measure of the remote interior angle at A.

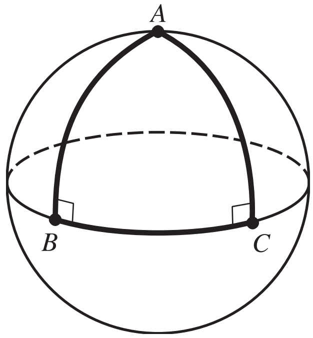

FIGURE 4.3: A spherical triangle

It is instructive to follow the steps of Euclid's proof on the sphere. All the steps work there except for the one that Euclid did not justify. If the construction in the proof of Theorem 4.1.2 is carried out on the sphere, starting with the triangle pictured, the point F constructed is in the interior of $\angle ACD$ if and only if $\mu(\angle BAC) < 90°$ (Exercise 2).

The second theorem of the section asserts that perpendicular lines exist and are unique. This result will be a fundamentally important tool in our development of neutral geometry. The existence part of the theorem is closely related to Euclid's Proposition 12, although Euclid meant something quite different by his proposition. Euclid was asserting that the perpendicular could be constructed using straightedge and compass; we are simply saying that the perpendicular line exists. The proof of uniqueness is based on the observation that the Exterior Angle Theorem prohibits triangles from having two right angles. Notice that, in the spherical triangle shown above, both $\overleftrightarrow{AB}$ and $\overleftrightarrow{AC}$ are lines through A that are perpendicular to $\overleftrightarrow{BC}$ so uniqueness of perpendiculars fails in spherical geometry.

Theorem 4.1.3 (Existence and Uniqueness of Perpendiculars). *For every line ℓ and for every point P, there exists a unique line m such that P lies on m and $m \perp \ell$.*

While we can prove both existence and uniqueness of perpendicular lines in neutral geometry, we are not able to do the same for parallel lines. As we will see, it is only possible to prove existence of parallels in neutral geometry, not uniqueness.

Terminology. When we wish to apply Theorem 4.1.3 in our proofs, we will say "drop a perpendicular from P to ℓ." That statement is to be interpreted as an invocation of Theorem 4.1.3. By definition of perpendicular, there is a point F that lies on both ℓ and m; the point at which the perpendicular intersects ℓ is called the *foot* of the perpendicular.

Proof of theorem 4.1.3. Let ℓ be a line and let P be a point (hypothesis). The case in which P lies on ℓ is covered by Theorem 3.5.9, so we will assume that P is an external point for ℓ. We first show that there exists a line m such that $P \in m$ and $m \perp \ell$. There exist distinct points Q and Q' on ℓ (Ruler Postulate). There exists a point R, on the opposite side of ℓ from P, such that $\angle Q'QP \cong \angle Q'QR$ (Angle Construction Postulate). Choose a point P' on $\overrightarrow{QR}$ such that $\overline{QP} \cong \overline{QP'}$ (Point Construction Postulate).

Let $m = \overleftrightarrow{PP'}$; the proof of existence will be complete if we show that $m \perp \ell$. There exists a point $F \in \overline{PP'} \cap \ell$ (Plane Separation Postulate). In case $Q = F$, $\angle Q'FP$ and $\angle Q'FP'$ form a linear pair and thus are supplements (Linear Pair Theorem). Since they are also congruent (earlier statement), they must both be right angles and so $m \perp \ell$. In case F lies on $\overrightarrow{QQ'}$, we have $\angle PQF = \angle PQQ'$ and $\angle P'QF = \angle P'QQ'$ and so $\angle PQF \cong \angle P'QF$ (choice of P'). Hence $\triangle FQP \cong \triangle FQP'$ (SAS), and $\angle QFP \cong \angle QFP'$ (definition of congruent triangles). Since $\angle QFP$ and $\angle QFP'$ form a linear pair they must both be right angles by the same argument as before and we conclude that $m \perp \ell$ in this case as well. The final case is that in which F lies on the ray opposite to $\overrightarrow{QQ'}$. In that case $\angle PQF$ and $\angle PQQ'$ are supplements as are $\angle P'QF$ and $\angle P'QQ'$. Again we have $\triangle FQP \cong \triangle FQP'$ (SAS) and the proof is completed as before. This completes the proof of existence.

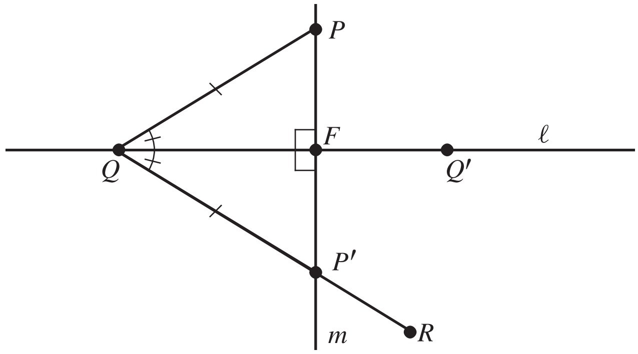

FIGURE 4.4: Proof of existence of perpendiculars

We now turn our attention to the proof of uniqueness. Suppose there exists a line m', different from m, such that P lies on m' and $m' \perp \ell$ (RAA hypothesis). Let Q' be the point at which ℓ and m' intersect. Then $Q' \neq Q$ (Theorem 3.1.7). Now $\triangle PQQ'$ has an exterior angle at Q' that measures 90° and an interior angle at Q of the same measure. This contradicts the Exterior Angle Theorem, which says that the measure of an exterior angle must be strictly greater than the measure of either remote interior angle. Thus we must reject the RAA hypothesis and conclude that no such line m' exists. □

More on terminology. The proof above lays out a step-by-step process in which we apply the various postulates we have assumed, along with the theorems we have already proved, to verify that the perpendicular line must exist. When we present such a proof, we often say that we have "constructed" the perpendicular. What we mean by this is simply that we have shown that that existence of the perpendicular follows from the axioms. This use of the word "construction" is quite different from what Euclid meant by a construction. Constructions in the sense of Euclid will be studied in Chapter 9.

In the proof above, we proved uniqueness by showing that the existence of a second perpendicular line would lead to a contradiction. A slightly different strategy for proving uniqueness is to show that any line through P that is perpendicular to ℓ must in fact be equal to m. Here is how that proof would be formulated: Assume m' is a line such that P lies on m' and $m' \perp \ell$. Let Q' be the point at which m' intersects ℓ (which exists by definition of perpendicular). If $Q' \neq Q$, we would have a triangle $\triangle PQQ'$ whose properties contradict the Exterior Angle Theorem, just as before. Hence it must be the case that $Q' = Q$ and therefore $m = m'$ by the Incidence Postulate. While the overall strategy for this second formulation of the proof is different, the essential proof is exactly the same as in the first formulation. A common error is to think that the Angle Construction Postulate implies uniqueness directly, but that is only the case if P lies on ℓ.

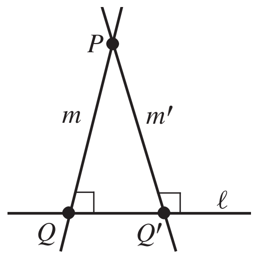

FIGURE 4.5: Two perpendiculars from *P*

EXERCISES 4.1

1. Prove: If one interior angle of a triangle is right or obtuse, then both the other interior angles are acute.

2. Let $\triangle ABC$ be the spherical triangle shown in Figure 4.3. Perform the construction in the proof of Theorem 4.1.2 on the sphere, starting with this triangle. Convince yourself that the point F constructed is in the interior of $\angle ACD$ if and only if $\mu(\angle BAC) < 90°$. Draw diagrams illustrating both possibilities.

4.2 TRIANGLE CONGRUENCE CONDITIONS

The statement that two triangles are congruent means that the three interior angles of the first triangle are congruent to the corresponding angles in the second and that the three sides of the first triangle are congruent to the corresponding sides of the second. The Side-Angle-Side Postulate indicates that it is sometimes possible to conclude all six of these congruences from only three of them. SAS is just the first of several similar results, which are known as *triangle congruence conditions*. In this section we build on the Side-Angle-Side Postulate to prove the other familiar triangle congruence conditions. We begin with the Angle-Side-Angle triangle congruence condition. It is one half of Euclid's Proposition 26; the other half (Angle-Angle-Side) will be left as an exercise.

Theorem 4.2.1 (ASA). *If two angles and the included side of one triangle are congruent to the corresponding parts of a second triangle, then the two triangles are congruent.*

Restatement. *If $\triangle ABC$ and $\triangle DEF$ are two triangles such that $\angle CAB \cong \angle FDE$, $\overline{AB} \cong \overline{DE}$, and $\angle ABC \cong \angle DEF$, then $\triangle ABC \cong \triangle DEF$.*

Proof. Let $\triangle ABC$ and $\triangle DEF$ be two triangles such that $\angle CAB \cong \angle FDE$, $\overline{AB} \cong \overline{DE}$, and $\angle ABC \cong \angle DEF$ (hypothesis). We must show that $\triangle ABC \cong \triangle DEF$.

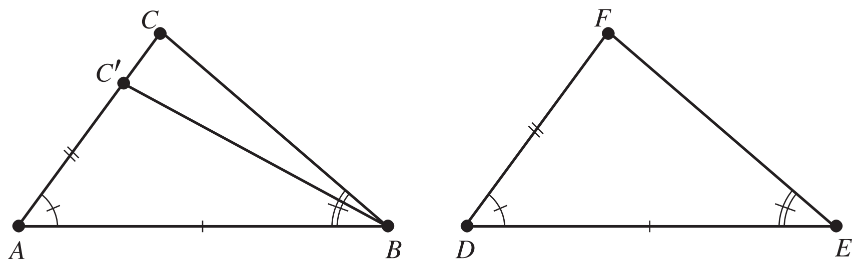

FIGURE 4.6: One possible location for C' in proof of ASA

There exists a point C' on $\overrightarrow{AC}$ such that $\overline{AC'} \cong \overline{DF}$ (Point Construction Postulate). Now $\triangle ABC' \cong \triangle DEF$ (SAS) and so $\angle ABC' \cong \angle DEF$ (definition of congruent triangles). Since $\angle ABC \cong \angle DEF$ (hypothesis), we can conclude that $\angle ABC \cong \angle ABC'$. Hence $\overrightarrow{BC} = \overrightarrow{BC'}$ (Protractor Postulate, Part 3). But $\overrightarrow{BC}$ can only intersect $\overleftrightarrow{AC}$ in at most one point (Theorem 3.1.7), so $C = C'$ and the proof is complete. □

Angle-Side-Angle can be used to prove the converse to the Isosceles Triangle Theorem, which is Euclid's Proposition 6.

Theorem 4.2.2 (Converse to the Isosceles Triangle Theorem). *If $\triangle ABC$ is a triangle such that $\angle ABC \cong \angle ACB$, then $\overline{AB} \cong \overline{AC}$.*

Proof. Exercise 1. □

Theorem 4.2.3 (AAS). *If $\triangle ABC$ and $\triangle DEF$ are two triangles such that $\angle ABC \cong \angle DEF$, $\angle BCA \cong \angle EFD$, and $\overline{AC} \cong \overline{DF}$, then $\triangle ABC \cong \triangle DEF$.*

Proof. Exercise 2. □

The two conditions ASA and AAS show that any two angles plus a side determine the entire triangle. Two sides plus one angle may or may not determine the triangle. If the angle is included between the two sides, then SAS implies that the triangle is determined. If the angle is not included between the sides, then the triangle is not completely determined and therefore ASS is not a valid congruence condition. The failure of Angle-Side-Side is illustrated in Figure 4.7. Even though the figure makes it clear that ASS is not a valid triangle congruence condition, two triangles that satisfy the hypotheses of Angle-Side-Side must be closely related—the relationship between them is specified in Exercise 3.

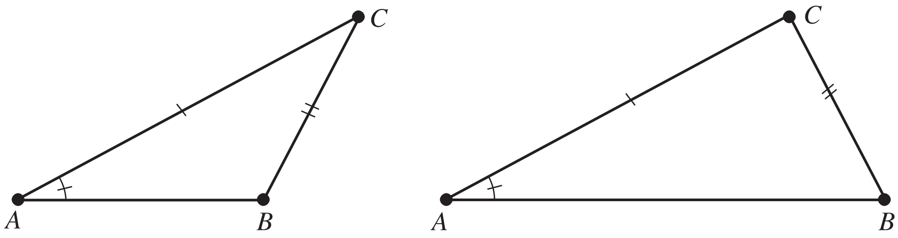

FIGURE 4.7: There is no Angle-Side-Side Theorem

There is one significant special case in which Angle-Side-Side does hold; that is the case in which the given angle is a right angle. The theorem is known as the Hypotenuse-Leg Theorem.

Definition 4.2.4. A triangle is a *right triangle* if one of the interior angles is a right angle. The side opposite the right angle is called the *hypotenuse* and the two sides adjacent to the right angle are called *legs*.

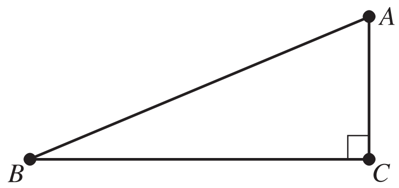

FIGURE 4.8: $\triangle ABC$ is a right triangle; $\overline{AB}$ is the hypotenuse; $\overline{AC}$ and $\overline{BC}$ are legs

Theorem 4.2.5 (Hypotenuse-Leg Theorem). *If the hypotenuse and one leg of a right triangle are congruent to the hypotenuse and a leg of a second right triangle, then the two triangles are congruent.*

Restatement. *If $\triangle ABC$ and $\triangle DEF$ are two right triangles with right angles at the vertices C and F, respectively, $\overline{AB} \cong \overline{DE}$, and $\overline{BC} \cong \overline{EF}$, then $\triangle ABC \cong \triangle DEF$.*

Proof. Exercise 4. □

Side-Side-Side is a also a valid triangle congruence condition. Its proof requires the following result, which is roughly equivalent to Euclid's Proposition 7.

Theorem 4.2.6. *If $\triangle ABC$ is a triangle, $\overline{DE}$ is a segment such that $\overline{DE} \cong \overline{AB}$, and H is a half-plane bounded by $\overleftrightarrow{DE}$, then there is a unique point $F \in H$ such that $\triangle DEF \cong \triangle ABC$.*

Proof. Exercise 5. □

Theorem 4.2.7 (SSS). *If $\triangle ABC$ and $\triangle DEF$ are two triangles such that $\overline{AB} \cong \overline{DE}$, $\overline{BC} \cong \overline{EF}$, and $\overline{CA} \cong \overline{FD}$, then $\triangle ABC \cong \triangle DEF$.*

Proof. Let $\triangle ABC$ and $\triangle DEF$ be two triangles such that $\overline{AB} \cong \overline{DE}$, $\overline{BC} \cong \overline{EF}$ and $\overline{CA} \cong \overline{FD}$ (hypothesis). We must prove that $\triangle ABC \cong \triangle DEF$.

There exists a point G, on the opposite of $\overleftrightarrow{AB}$ from C, such that $\triangle ABG \cong \triangle DEF$ (Theorem 4.2.6). Since C and G are on opposite sides of $\overleftrightarrow{AB}$, there is a point H such that H is between C and G and H lies on $\overleftrightarrow{AB}$ (Plane Separation Postulate). It is possible that $H = A$ or $H = B$. If not, then exactly one of the following will hold: $H * A * B$, or $A * H * B$, or $A * B * H$ (Corollary 3.2.19). Thus there are five possibilities for the location of H relative to A and B. We will consider the various cases separately.

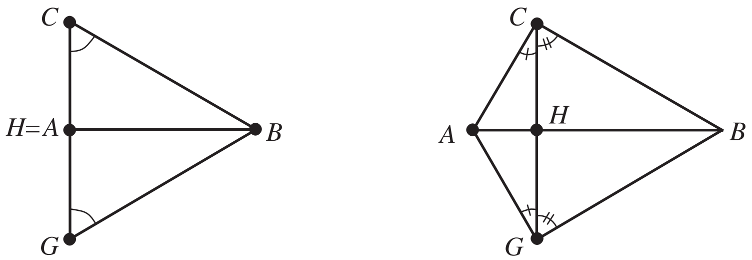

FIGURE 4.9: Case 1, $H = A$, and Case 3, $A * H * B$

Case 1. $H = A$. In this case $\angle ACB = \angle GCB$ and $\angle AGB = \angle CGB$. Since $BC = BG$, we have that $\angle GCB \cong \angle CGB$ (Isosceles Triangle Theorem). Thus $\triangle ABC \cong \triangle ABG$ (SAS).
Case 2. $H = B$. The proof of this case is similar to the proof of Case 1.
Case 3. $A * H * B$. Since H is between A and B, H is in the interior of $\angle ACB$ and in the interior of $\angle AGB$ (Theorem 3.3.10). Thus $\mu(\angle ACB) = \mu(\angle ACH) + \mu(\angle HCB)$ and $\mu(\angle AGB) = \mu(\angle AGH) + \mu(\angle HGB)$ (Angle Addition Postulate). But $\mu(\angle ACH) = \mu(\angle ACG) = \mu(\angle AGC) = \mu(\angle AGH)$ (Isosceles Triangle Theorem). In the same way, $\mu(\angle HCB) = \mu(\angle HGB)$. Thus $\angle ACB \cong \angle AGB$ (addition) and so $\triangle ABC \cong \triangle ABG$ (SAS).

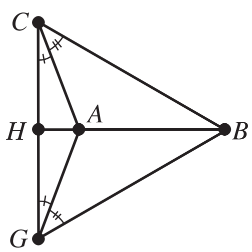

FIGURE 4.10: Case 4, $H * A * B$

Case 4. $H * A * B$. The proof of this case is very much like the proof of Case 3. This time A is in the interior of $\angle HCB$ and $\angle HGB$ (see Figure 4.10). Thus $\mu(\angle ACB) = \mu(\angle HCB) - \mu(\angle HCA)$ and $\mu(\angle AGB) = \mu(\angle HGB) - \mu(\angle HGA)$ (Protractor Postulate, Part 4). But $\mu(\angle ACH) = \mu(\angle AGH)$ and $\mu(\angle HCB) = \mu(\angle HCB)$ (Isosceles Triangle Theorem). Thus $\angle ACB \cong \angle AGB$ (subtraction) and so $\triangle ABC \cong \triangle ABG$ (SAS).
Case 5. $A * B * H$. The proof of this case is similar to the proof of Case 4.

In each of the five cases we have proved that $\triangle ABC \cong \triangle ABG$. But $\triangle ABG \cong \triangle DEF$, so the proof is complete. □

Side-Side-Side is Euclid's Proposition 8. Euclid's proof is somewhat clumsy, relying on a rather convoluted lemma (Proposition 7). Heath attributes the more elegant proof, above, to Philo and Proclus (see [22], page 263). SSS can also be proved as a corollary of the Hinge Theorem (see Exercise 4.3.5), but that proof is not considered to be as elegant as the proof given here.

Before leaving the subject of triangle congruence conditions we should mention that in neutral geometry the three angles of a triangle do not determine the triangle. This may seem obvious to you since you remember from high school that similar triangles are usually not congruent. However, that is a special situation in Euclidean geometry. One of the surprising results we will prove in hyperbolic geometry is that (in that context) Angle-Angle-Angle is a valid triangle congruence condition!

EXERCISES 4.2

1. Prove the Converse to the Isosceles Triangle Theorem (Theorem 4.2.2).
2. Prove the Angle-Angle-Side Triangle Congruence Condition (Theorem 4.2.3).
3. Suppose $\triangle ABC$ and $\triangle DEF$ are two triangles such that $\angle BAC \cong \angle EDF$, $AC = DF$, and $CB = FE$ (the hypotheses of ASS). Prove that either $\angle ABC$ and $\angle DEF$ are congruent or they are supplements.
4. Prove the Hypotenuse-Leg Theorem (Theorem 4.2.5).
5. Prove that it is possible to construct a congruent copy of a triangle on a given base (Theorem 4.2.6).

4.3 THREE INEQUALITIES FOR TRIANGLES

The Exterior Angle Theorem gives one inequality that is always satisfied by the measures of the angles of a triangle. In this section we prove three additional inequalities that will be useful in our study of triangles. The first of these theorems, the Scalene Inequality, extends the Isosceles Triangle Theorem and its converse. It combines Euclid's Propositions 18 and 19. The word *scalene* means "unequal" or "uneven." A *scalene triangle* is a triangle that has sides of three different lengths.

Theorem 4.3.1 (Scalene Inequality). *In any triangle, the greater side lies opposite the greater angle and the greater angle lies opposite the greater side.*

Restatement. *Let $\triangle ABC$ be a triangle. Then $AB > BC$ if and only if $\mu(\angle ACB) > \mu(\angle BAC)$.*

Proof. Let A, B, and C be three noncollinear points (hypothesis). We will first assume the hypothesis $AB > BC$ and prove that $\mu(\angle ACB) > \mu(\angle BAC)$. Since $AB > BC$, there exists a point D between A and B such that $\overline{BD} \cong \overline{BC}$ (Ruler Postulate).

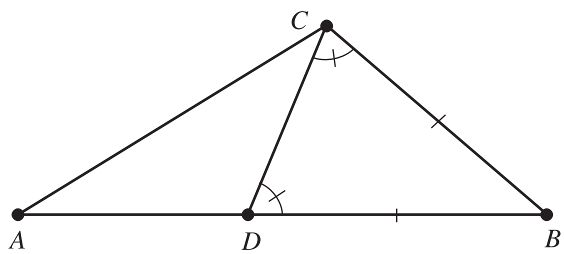

FIGURE 4.11: The greater side lies opposite greater angle

Now $\mu(\angle ACB) > \mu(\angle DCB)$ (Protractor Postulate, Part 4, and Theorem 3.3.10) and $\angle DCB \cong \angle CDB$ (Isosceles Triangle Theorem). But $\angle CDB$ is an exterior angle for

$\triangle ADC$ (see Figure 4.11), so $\mu(\angle CDB) > \mu(\angle CAB)$ (Exterior Angle Theorem). The conclusion follows from those inequalities.

The proof of the converse is left as an exercise (Exercise 1). □

The second inequality is the familiar Triangle Inequality. It is Euclid's Proposition 20.

Theorem 4.3.2 (Triangle Inequality). *If A, B, and C are three noncollinear points, then* $AC < AB + BC$.

Proof. Exercise 2. □

The third inequality is Euclid's Proposition 24. It is known as the Hinge Theorem. The reason for the name is the fact that the lengths of the two sides of the triangle are fixed but the angle is allowed to vary. This means that the triangle can open and close, much like a door hinge. The theorem is a generalization of SAS and is sometimes called the SAS Inequality.

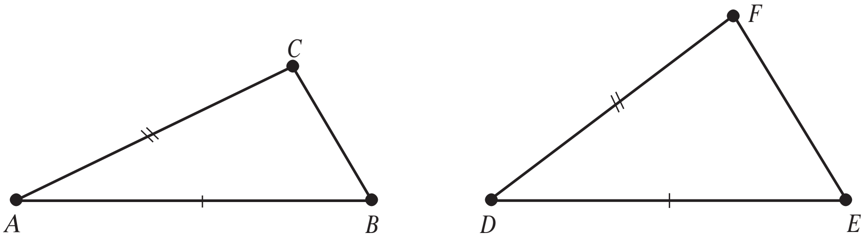

FIGURE 4.12: Hinge Theorem: If $\mu(\angle BAC) < \mu(\angle EDF)$, then $BC < EF$

Theorem 4.3.3 (Hinge Theorem). *If* $\triangle ABC$ *and* $\triangle DEF$ *are two triangles such that* $AB = DE$, $AC = DF$, *and* $\mu(\angle BAC) < \mu(\angle EDF)$, *then* $BC < EF$.

Proof. Let $\triangle ABC$ and $\triangle DEF$ be two triangles such that $AB = DE$, $AC = DF$, and $\mu(\angle BAC) < \mu(\angle EDF)$. We must show that $BC < EF$. Find a point G, on the same side of $\overleftrightarrow{AB}$ as C, such that $\triangle ABG \cong \triangle DEF$ (Theorem 4.2.6). The proof will be complete if we show that $BG > BC$.

Since C is in the interior of $\angle BAG$ (Theorem 3.4.5), $\overrightarrow{AC}$ must intersect $\overline{BG}$ in a point J (Crossbar Theorem). If $J = C$, then C is between B and G and the conclusion $BG > BC$ follows immediately. Hence we may assume for the remainder of the proof that $J \neq C$, which means that C does not lie on $\overleftrightarrow{BG}$ (Theorem 3.1.7).

Let $\overrightarrow{AH}$ be the bisector of $\angle CAG$ (Theorem 3.4.7). Then H is in the interior of $\angle CAG = \angle JAG$ (definition of angle bisector), so $\overrightarrow{AH}$ must intersect $\overline{JG}$ (Crossbar Theorem). In order to simplify the notation, we will assume that H lies on $\overline{JG}$. Since $\overline{JG} \subset \overline{BG}$, H is between B and G. Note that $\triangle AHG \cong \triangle AHC$ (SAS), so $HG = HC$. Hence $BG = BH + HG = BH + HC$. Since C does not lie on $\overleftrightarrow{BH}$, the Triangle Inequality yields $BC < BH + HC$, which completes the proof. □

Euclid's proof of the Hinge Theorem is quite complicated and so we have given a different proof.[1] One reason Euclid's proof is so complicated is the fact that it has three cases that must be considered. In proofs that involved multiple cases, Euclid followed the

[1] The proof included here is the one Heath calls a "modern" proof of the theorem [22, page 298].

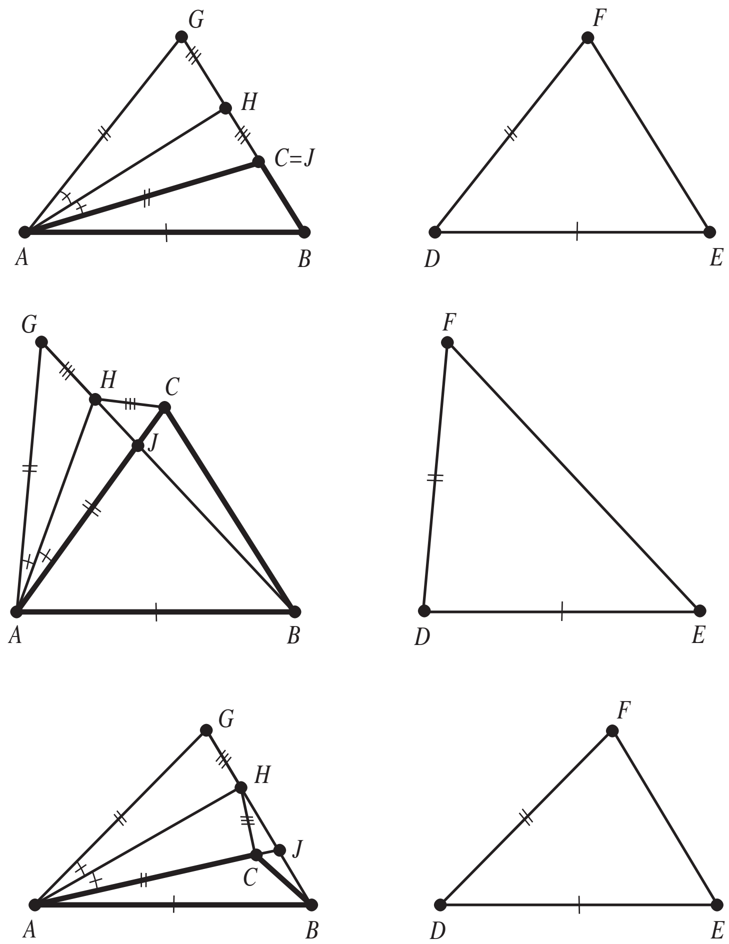

FIGURE 4.13: Three possible diagrams of the proof of the Hinge Theorem

practice of including the details of only one case, usually the most difficult one. As a result, his written proof cannot really be considered complete by modern standards. It would be easy to fill in the omitted details, but the result is a rather intricate and complicated proof. The nice thing about the proof we gave above is the fact that it is only necessary to consider two different cases ($C = J$, $C \neq J$). It should be noted, however, that three different diagrams are possible: see Figure 4.13.

The Scalene Inequality allows us to prove that the perpendicular is the shortest line segment joining an external point to a line.

Theorem 4.3.4. *Let ℓ be a line, let P be an external point, and let F be the foot of the perpendicular from P to ℓ. If R is any point on ℓ that is different from F, then $PR > PF$.*

Proof. Exercise 7 (see Figure 4.14.) □

The theorem motivates the following definition.

Definition 4.3.5. If ℓ is a line and P is a point, the *distance from P to ℓ*, denoted $d(P, \ell)$, is defined to be the distance from P to the foot of the perpendicular from P to ℓ.

We can use that definition to characterize the points that lie on an angle bisector.

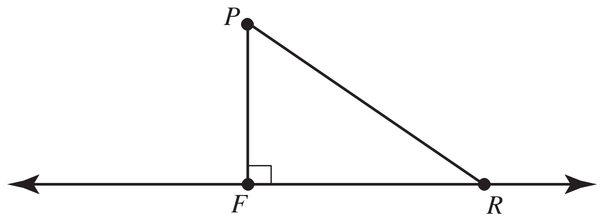

FIGURE 4.14: $PR > PF$

Theorem 4.3.6 (Pointwise Characterization of Angle Bisector). *Let A, B, and C be three noncollinear points and let P be a point in the interior of* $\angle BAC$. *Then P lies on the angle bisector of* $\angle BAC$ *if and only if* $d(P, \overleftrightarrow{AB}) = d(P, \overleftrightarrow{AC})$.

Proof. Exercise 8 (see Figure 4.15). □

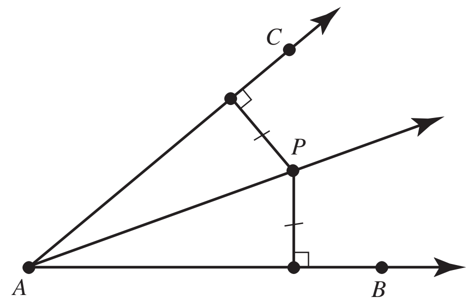

FIGURE 4.15: Points on the angle bisector

There is a similar characterization of points that lie on a perpendicular bisector.

Theorem 4.3.7 (Pointwise Characterization of Perpendicular Bisector). *Let A and B be distinct points. A point P lies on the perpendicular bisector of* $\overline{AB}$ *if and only if* $PA = PB$.

Proof. Exercise 9 (see Figure 4.16). □

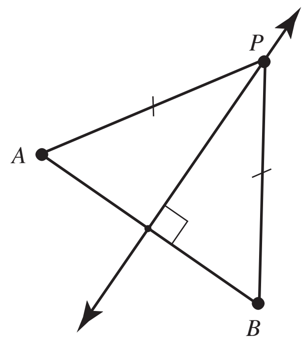

FIGURE 4.16: Points on the perpendicular bisector

We end this section with an application of the Triangle Inequality that is very different from anything in Euclid. We will prove that the function that measures distances is continuous in a certain technical sense. This theorem will allow us to bring results from calculus and real analysis (such as the Intermediate Value Theorem) to bear on geometry. The theorem is not used until Chapters 7 and 8, so most readers will probably want to omit it for now and refer back to it only when it is needed.

Setting for the Continuity of Distance Theorem. Let A, B, and C be three noncollinear points. Let $d = AB$. For each $x \in [0, d]$ there exists a unique point $D_x \in \overline{AB}$ such that $AD_x = x$ (Ruler Postulate). Define a function $f : [0, d] \to [0, \infty)$ by $f(x) = CD_x$.

Theorem 4.3.8 (Continuity of Distance). *The function f defined in the previous paragraph is continuous.*

Proof. Let A, B, C, and f be as in the setting for the theorem. We will show that f is continuous at x for $0 < x < d$. The proof of continuity at an endpoint of $[0, d]$ is similar. Let $\epsilon > 0$ be given. We must show that there exists $\delta > 0$ such that if $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$ (definition of continuous at x). We claim that $\delta = \epsilon$ works. Suppose y is a number in $[0, d]$ such that $|x - y| < \epsilon$. By the Triangle Inequality, $CD_x < CD_y + D_xD_y < CD_y + \epsilon$. Applying the Triangle Inequality again gives $CD_y < CD_x + D_xD_y < CD_x + \epsilon$. Combining the two inequalities yields $CD_x - \epsilon < CD_y < CD_x + \epsilon$. Therefore $|f(x) - f(y)| = |CD_x - CD_y| < \epsilon$ as required. □

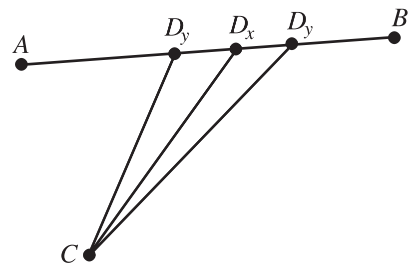

FIGURE 4.17: Two possible locations for D_y in the proof of continuity

EXERCISES 4.3

1. Complete the proof of the Scalene Inequality (Theorem 4.3.1).
2. Prove the Triangle Inequality (Theorem 4.3.2).
3. Prove the following result: *A, B, and C are three points such that $AB + BC = AC$, then A, B, and C are collinear.* (It follows that the assumption in the definition of between that A, B, and C are collinear is redundant.)
4. Prove: If A, B, and C are any three points (collinear or not), then $AB + BC \geq AC$.
5. Use the Hinge Theorem to prove SSS.
6. Prove that the hypotenuse is always the longest side of a right triangle.
7. Prove that the shortest distance from a point to a line is measured along the perpendicular (Theorem 4.3.4).
8. Prove the Pointwise Characterization of Angle Bisectors (Theorem 4.3.6).
9. Prove the Pointwise Characterization of Perpendicular Bisectors (Theorem 4.3.7).

4.4 THE ALTERNATE INTERIOR ANGLES THEOREM

We come now to our first theorem about parallel lines. The Alternate Interior Angles Theorem is one of the major theorems of neutral geometry. It is also Euclid's Proposition 27. Before we can state the theorem we need some definitions and notation.

Definition 4.4.1. Let ℓ and ℓ' be two distinct lines. A third line t is called a *transversal* for ℓ and ℓ' if t intersects ℓ in one point B and t intersects ℓ' in one point B' with $B' \neq B$. Notice that there are four angles with vertex B that are formed by rays on t and ℓ and there are four angles with vertex B' that are formed by rays on t and ℓ'. We will give names to some of these eight angles. In order to do so, choose points A and C on ℓ and A' and C' on ℓ' so that $A * B * C$, $A' * B' * C'$, and A and A' are on the same side of t. (Such points exist by the Ruler and Plane Separation Postulates.) The four angles $\angle ABB'$, $\angle A'B'B$, $\angle CBB'$, and $\angle C'B'B$ are called *interior angles* for ℓ and ℓ' with transversal t. The two pairs $\{\angle ABB', \angle BB'C'\}$ and $\{\angle A'B'B, \angle B'BC\}$ are called *alternate interior angles*. See Figure 4.18.

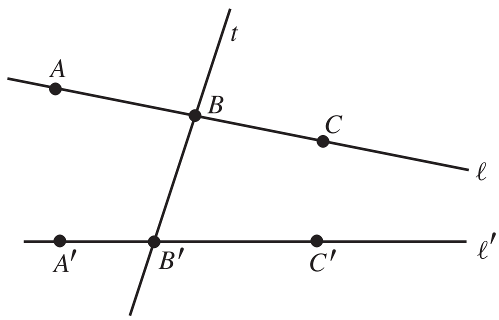

FIGURE 4.18: Definition of transversal

When t is a transversal for lines ℓ and ℓ', we say that "ℓ and ℓ' are *cut* by transversal t." This statement means that t intersects both lines and that it does so at distinct points.

Theorem 4.4.2 (Alternate Interior Angles Theorem). *If ℓ and ℓ' are two lines cut by a transversal t in such a way that a pair of alternate interior angles is congruent, then ℓ is parallel to ℓ'.*

Proof. Let ℓ and ℓ' be two lines cut by transversal t such that a pair of alternate interior angles is congruent (hypothesis). Choose points A, B, C and A', B', C' as in the definition of transversal above. Let us say that $\angle A'B'B \cong \angle B'BC$ (hypothesis). We must prove that $\ell \parallel \ell'$.

Suppose there exists a point D such that D lies on both ℓ and ℓ' (RAA hypothesis). If D lies on the same side of t as C, then $\angle A'B'B$ is an exterior angle for $\triangle BB'D$ while $\angle B'BC$ is a remote interior angle for that same triangle (definitions). This contradicts the Exterior Angles Theorem (Figure 4.19).

In case D lies on the same side of t as A, then $\angle B'BC$ is an exterior angle and $\angle A'B'B$ is a remote interior angle for $\triangle BB'D$ and we again reach a contradiction. Since D must lie on one of the two sides of t (Plane Separation Postulate), we are forced to reject the RAA hypothesis and the theorem is proved. □

Notice that one pair of alternate interior angles is congruent if and only if the other one is. This follows from the fact that supplements of congruent angles are congruent, which was an exercise in the previous chapter.

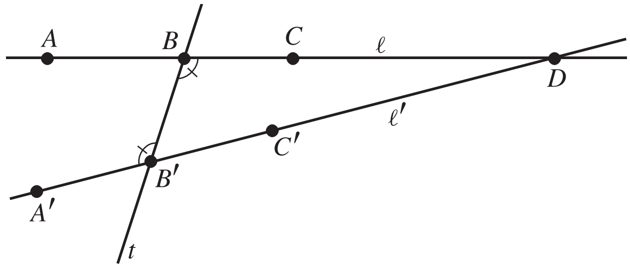

FIGURE 4.19: Impossible alternate interior angles

It is also common to state the theorem in terms of corresponding angles. This version is Euclid's Proposition 28.

Definition 4.4.3. Keep the notation used in the definition of interior angles and choose a point B'' on t such that $B * B' * B''$ (Ruler Postulate). The angles $\{\angle B'BC, \angle B''B'C'\}$ are called *corresponding angles*. There are three other pairs of corresponding angles that are defined in the obvious way.

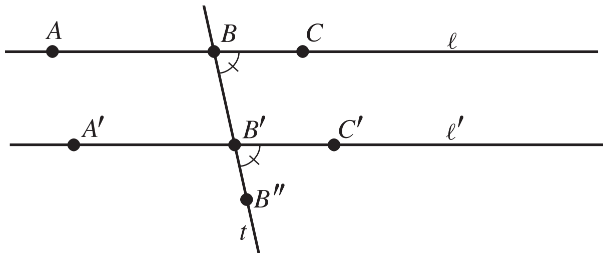

FIGURE 4.20: Congruent corresponding angles

Corollary 4.4.4 (Corresponding Angles Theorem). *If ℓ and ℓ' are lines cut by a transversal t in such a way that two corresponding angles are congruent, then ℓ is parallel to ℓ'.*

Proof. Exercise 1. □

The Alternate Interior Angles Theorem has several other interesting and important corollaries. The first is a partial converse to Euclid's Fifth Postulate.

Corollary 4.4.5. *If ℓ and ℓ' are lines cut by a transversal t in such a way that two nonalternating interior angles on the same side of t are supplements, then ℓ is parallel to ℓ'.*

Proof. Exercise 2. □

The next corollary, which asserts that parallel lines exist, is one of the most important consequences of the Alternate Interior Angles Theorem. The construction in the proof is simple, but it is one that we will use over and over in the remainder of the course. In fact, most applications of the corollary will use the construction from the proof rather than the statement of the corollary itself. Be sure to notice that there is no claim that the parallel constructed in the proof is unique. As we will see later, the uniqueness of parallels cannot be proved in neutral geometry. The existence of parallels is Euclid's Proposition 31; as usual Euclid was really asserting that the parallel could be constructed with straightedge and compass.

Corollary 4.4.6 (Existence of Parallels). *If ℓ is a line and P is an external point, then there is a line m such that P lies on m and m is parallel to ℓ.*

Proof. Let ℓ be a line and let P be an external point (hypothesis). Drop a perpendicular from P to ℓ (Theorem 4.1.3). Call the foot of that perpendicular Q and let $t = \overleftrightarrow{PQ}$. Next construct a line m through P that is perpendicular to t (Theorem 3.5.9). Then $\ell \parallel m$ by the Alternate Interior Angles Theorem (all the interior angles are right angles). □

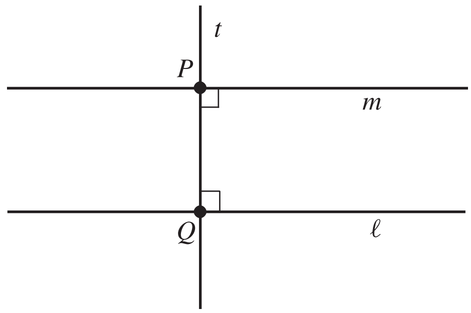

FIGURE 4.21: Proof of the existence of parallels

When Corollary 4.4.6 is applied, it is often the construction itself that is needed, rather than the result of that construction. For future reference, let us name this construction *the double perpendicular construction of a parallel line.*

The Double Perpendicular Construction. *Given a line ℓ and an external point P, drop a perpendicular t from P to ℓ. Let m be the line through P that is perpendicular to t. Then P lies on both t and m, m is parallel to ℓ, and t is a transversal for ℓ and m that is perpendicular to both ℓ and m.*

The existence of parallels implies, in particular, that the Elliptic Parallel Postulate is not consistent with the axioms of neutral geometry. Thus we can state the following corollary.

Corollary 4.4.7. *The Elliptic Parallel Postulate is false in any model for neutral geometry.*

The following special case of the Alternate Interior Angles Theorem is often useful.

Corollary 4.4.8. *If ℓ, m, and n are three lines such that $m \perp \ell$ and $n \perp \ell$, then either $m = n$ or $m \parallel n$.*

Proof. Exercise 3. □

EXERCISES 4.4

1. Prove the Corresponding Angles Theorem (Corollary 4.4.4).
2. Prove Corollary 4.4.5.
3. Prove Corollary 4.4.8.

4.5 THE SACCHERI-LEGENDRE THEOREM

The next theorem tells us about angle sums for triangles. The theorem is named for two mathematicians who contributed to our understanding of the theorem's place in neutral geometry. It is our first major departure from Euclid's propositions and also the first time we need to use one of the deep properties of real numbers.

Definition 4.5.1. Let A, B, and C be three noncollinear points. The *angle sum* for $\triangle ABC$ is the sum of the measures of the three interior angles of $\triangle ABC$. More specifically, the angle sum is defined by the equation

$$\sigma(\triangle ABC) = \mu(\angle CAB) + \mu(\angle ABC) + \mu(\angle BCA).$$

It is clear from the definition that congruent triangles have equal angle sums.

We will prove that the angle sum of any triangle in neutral geometry is less than or equal to $180°$. The theorem is not one of the familiar theorems from high school geometry nor is it one of Euclid's propositions. Euclid proves the well known "fact" that the angle sum of any triangle is exactly equal to $180°$. His proof uses his fifth postulate, so it has no place in neutral geometry. We will see later that there are models for neutral geometry in which triangles have angle sums strictly less than $180°$ and that angle sums are one of the major distinguishing characteristics for Euclidean and non-Euclidean geometries. What the Saccheri-Legendre Theorem rules out is angle sums that are strictly greater than $180°$. Be sure to notice that triangles on the sphere $\mathbb{S}^2$ have angle sums that are strictly greater than $180°$ (see Figure 4.3) and so the theorem must rely in a substantial way on the axioms of neutral geometry.

Theorem 4.5.2 (Saccheri-Legendre Theorem). *If $\triangle ABC$ is any triangle, then $\sigma(\triangle ABC) \leqslant 180°$.*

The proof of the theorem relies on several lemmas. The first tells us that the sum of the measures of any two angles in a triangle is always strictly less than $180°$ while the second tells us how angle sums add when a triangle is subdivided. The third (and most substantial) allows us to replace one triangle with another triangle that has the same angle sum but at least one angle that is smaller by a factor of two.

Lemma 4.5.3. *If $\triangle ABC$ is any triangle, then*

$$\mu(\angle CAB) + \mu(\angle ABC) < 180°.$$

Proof. Let $\triangle ABC$ be a triangle (hypothesis). We must prove that $\mu(\angle CAB) + \mu(\angle ABC) < 180°$.

Let D be a point on $\overleftrightarrow{AB}$ such that $A * B * D$. Then $\mu(\angle ABC) + \mu(\angle CBD) = 180°$ (Linear Pair Theorem). In addition, $\mu(\angle CAB) < \mu(\angle CBD$ (Exterior Angle Theorem). Therefore, $\mu(\angle CAB) + \mu(\angle ABC) < 180°$ (algebra). □

Lemma 4.5.4. *If $\triangle ABC$ is a triangle and E is a point in the interior of $\overline{BC}$, then*

$$\sigma(\triangle ABE) + \sigma(\triangle ECA) = \sigma(\triangle ABC) + 180°.$$

Proof. Let $\triangle ABC$ be a triangle and let E be a point in the interior of $\overline{BC}$ (hypothesis). We must prove that $\sigma(\triangle ABE) + \sigma(\triangle ECA) = \sigma(\triangle ABC) + 180°$.

By the definition of angle sum we have that $\sigma(\triangle ABE) + \sigma(\triangle ECA) = \mu(\angle EAB) + \mu(\angle ABE) + \mu(\angle BEA) + \mu(\angle CAE) + \mu(\angle ECA) + \mu(\angle AEC)$. Now E is in the interior of $\angle BAC$ (Theorem 3.3.10), so $\mu(\angle EAB) + \mu(\angle CAE) = \mu(\angle BAC)$ (Angle Addition Postulate). In addition, $\mu(\angle BEA) + \mu(\angle AEC) = 180°$ (Linear Pair Theorem). Therefore, $\sigma(\triangle ABE) + \sigma(\triangle ECA) = \mu(\angle BAC) + \mu(\angle ABE) + \mu(\angle ECA) + 180° = \sigma(\triangle ABC) + 180°$ (algebra and definition of angle sum). □

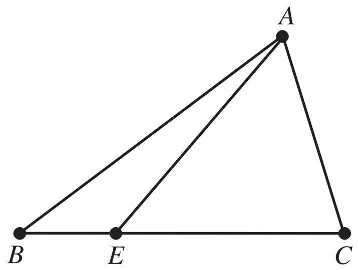

FIGURE 4.22: The triangle $\triangle ABC$ is subdivided into two smaller triangles

Lemma 4.5.5. *If A, B, and C are three noncollinear points, then there exists a point D that does not lie on $\overleftrightarrow{AB}$ such that $\sigma(\triangle ABD) = \sigma(\triangle ABC)$ and the angle measure of one of the interior angles in $\triangle ABD$ is less than or equal to $\frac{1}{2}\mu(\angle CAB)$.*

Proof. Let A, B, and C be three noncollinear points (hypothesis). Let E be the midpoint of $\overline{BC}$ (Theorem 3.2.22). Let D be a point on $\overrightarrow{AE}$ such that $A * E * D$ and $AE = ED$ (Ruler Postulate). We will show that $\sigma(\triangle ABD) = \sigma(\triangle ABC)$ and either $\mu(\angle BAD) \leqslant \frac{1}{2}\mu(\angle BAC)$ or $\mu(\angle ADB) \leqslant \frac{1}{2}\mu(\angle BAC)$.

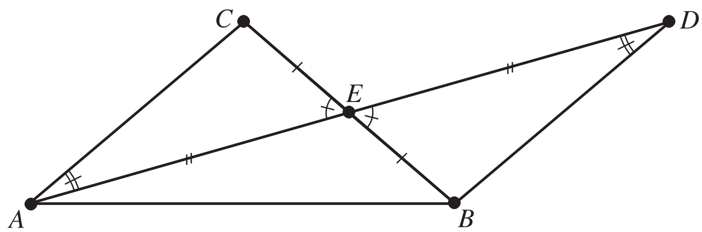

FIGURE 4.23: Proof of Lemma 4.5.5

Since $\angle AEC \cong \angle DEB$ (Vertical Angles Theorem), we see that $\triangle AEC \cong \triangle DEB$ (SAS) and so $\sigma(\triangle AEC) = \sigma(\triangle DEB)$. Applying Lemma 4.5.4 twice gives

$$\sigma(\triangle ABC) = \sigma(\triangle ABE) + \sigma(\triangle AEC) - 180^\circ$$

and

$$\sigma(\triangle ABD) = \sigma(\triangle ABE) + \sigma(\triangle DEB) - 180^\circ.$$

It follows that $\sigma(\triangle ABD) = \sigma(\triangle ABC)$.

Now $\mu(\angle BAE) + \mu(\angle EAC) = \mu(\angle BAC)$ (Protractor Postulate). Hence either $\mu(\angle BAE) \leqslant \frac{1}{2}\mu(\angle BAC)$ or $\mu(\angle EAC) \leqslant \frac{1}{2}\mu(\angle BAC)$ (algebra). Since $\mu(\angle EAC) = \mu(\angle ADB)$ (previous paragraph), the proof is complete. □

The proof that the lemmas imply the Saccheri-Legendre Theorem uses the Archimedean Property of Real Numbers. This is one of our basic assumptions about real numbers. It simply asserts that if a small positive quantity is doubled often enough, it will eventually grow larger than any fixed number. A precise statement is given in Appendix E, Axiom E.3.4

Proof of Theorem 4.5.2. Let $\triangle ABC$ be a triangle (hypothesis). Suppose $\sigma(\triangle ABC) > 180^\circ$ (RAA hypothesis). Let us say that $\sigma(\triangle ABC) = 180^\circ + \epsilon^\circ$, where ϵ is a positive real number. Choose a positive integer n large enough so that $2^n\epsilon^\circ > \mu(\angle CAB)$ (Archimedean Property of Real Numbers).

By Lemma 4.5.5, there is a triangle $\triangle A_1B_1C_1$ such that $\sigma(\triangle A_1B_1C_1) = \sigma(\triangle ABC)$ and one of the angles in $\triangle A_1B_1C_1$ has angle measure $\leqslant \frac{1}{2}\mu(\angle CAB)$. Applying the lemma a second time gives a triangle $\triangle A_2B_2C_2$ such that $\sigma(\triangle A_2B_2C_2) = \sigma(\triangle ABC)$ and one of the angles in $\triangle A_2B_2C_2$ has angle measure less than or equal to $\frac{1}{4}\mu(\angle CAB)$.

Applying Lemma 4.5.5 a total of n times yields a triangle $\triangle A_nB_nC_n$ such that $\sigma(\triangle A_nB_nC_n) = \sigma(\triangle ABC) = 180^\circ + \epsilon^\circ$ and one of the angles in $\triangle A_nB_nC_n$ has measure less than or equal to $\frac{1}{2^n}\mu(\angle CAB) < \epsilon^\circ$. Thus the sum of the measures of the other two angles is greater than 180° (algebra). But this contradicts Lemma 4.5.3. Hence we must reject the RAA hypothesis and the proof is complete. □

Here are two interesting corollaries of the Saccheri-Legendre Theorem. The first improves on the Exterior Angle Theorem. The second is a full converse to Euclid's Fifth Postulate. It is worth noting that, while Euclid's Fifth Postulate is not a theorem or an axiom of neutral geometry, its converse is a theorem.

Corollary 4.5.6. *The sum of the measures of two interior angles of a triangle is less than or equal to the measure of their remote exterior angle.*

Proof. Exercise 1. □

Corollary 4.5.7 (Converse to Euclid's Fifth Postulate). *Let ℓ and ℓ' be two lines cut by a transversal t. If ℓ and ℓ' meet on one side of t, then the sum of the measures of the two interior angles on that side of t is strictly less than* 180°.

Proof. Exercise 2. □

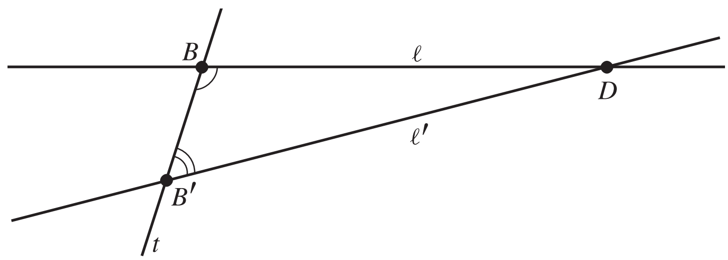

FIGURE 4.24: Converse to Euclid V: $\mu(\angle DBB') + \mu(\angle DB'B) < 180^\circ$

EXERCISES 4.5

1. Prove Corollary 4.5.6.
2. Prove the Converse to Euclid's Fifth Postulate (Corollary 4.5.7).

4.6 QUADRILATERALS

In our study of geometry we will want to work with four-sided figures (quadrilaterals) as well as three-sided ones (triangles). In particular, it will be important to understand angle sums for quadrilaterals. This section contains a brief introduction to quadrilaterals and their properties. We begin with some definitions.

Definition 4.6.1. Let A, B, C, and D be four points, no three of which are collinear. Suppose further that any two of the segments $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, and $\overline{DA}$ either have no point in common or have only an endpoint in common. If those conditions are satisfied, then the points A, B, C, and D determine a *quadrilateral*, which we will denote by $\square ABCD$. The quadrilateral is the union of the four segments $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, and $\overline{DA}$. The four segments are called the *sides* of the quadrilateral and the points A, B, C, and D are called the *vertices* of the quadrilateral. The sides $\overline{AB}$ and $\overline{CD}$ are called *opposite sides* of the quadrilateral as are the sides $\overline{BC}$ and $\overline{AD}$. Two quadrilaterals are *congruent* if there is a correspondence between their vertices so that all four corresponding sides are congruent and all four corresponding angles are congruent.

Notice that the order in which the vertices of a quadrilateral are listed is important. In general, $\square ABCD$ is a different quadrilateral from $\square ACBD$. In fact, there may not even be a quadrilateral $\square ACBD$ since the segments $\overline{AC}$ and $\overline{BD}$ may intersect at an interior point.

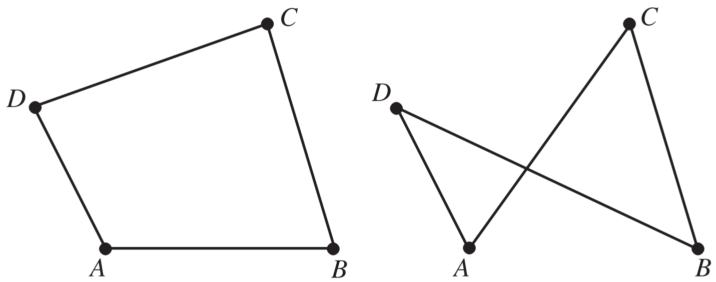

FIGURE 4.25: $\square ABCD$ is a quadrilateral, $\square ACBD$ is not

Definition 4.6.2. The *diagonals* of the quadrilateral $\square ABCD$ are the segments $\overline{AC}$ and $\overline{BD}$. The *angles* of the quadrilateral $\square ABCD$ are the angles $\angle ABC$, $\angle BCD$, $\angle CDA$, and $\angle DAB$. A quadrilateral is said to be *convex* if each vertex of the quadrilateral is contained in the interior of the angle formed by the other three vertices.

As Figure 4.26 suggests, it is possible to identify an interior for a quadrilateral and the interior is convex according to the earlier definition of convex set if and only if the quadrilateral is convex according to our new definition. Making a precise definition of interior of a quadrilateral involves complications that we prefer to avoid at present, however, so that approach is delayed until Chapter 7. Instead we use a definition that is easily formulated in terms of the concepts we have been studying in this chapter and the last.

Definition 4.6.3. If $\square ABCD$ is a convex quadrilateral, then the *angle sum* is defined by

$$\sigma(\square ABCD) = \mu(\angle ABC) + \mu(\angle BCD) + \mu(\angle CDA) + \mu(\angle DAB).$$

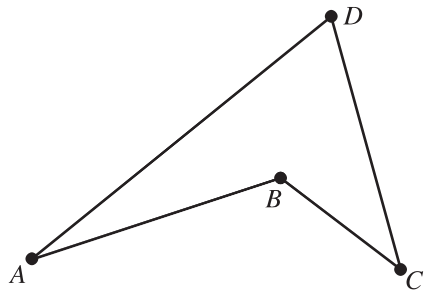

FIGURE 4.26: A nonconvex quadrilateral

We will not define angle sums for nonconvex quadrilaterals. The reason should be evident from Figure 4.26. What we intuitively want to call interior angles for the quadrilateral may not be angles at all, at least not according to the way we have defined angles. Each vertex does determine an angle, but one of those angles may be on the "exterior" of the quadrilateral. If we were to ignore that fact and simply use the formula to define an angle sum, we would have trouble proving theorems like the next proposition.

Theorem 4.6.4. *If $\square ABCD$ is a convex quadrilateral, then $\sigma(\square ABCD) \leqslant 360°$.*

Proof. Exercise 1. □

We end this section with several technical facts about convex quadrilaterals that will be needed in later chapters.

Definition 4.6.5. The quadrilateral $\square ABCD$ is called a *parallelogram* if $\overleftrightarrow{AB} \parallel \overleftrightarrow{CD}$ and $\overleftrightarrow{AD} \parallel \overleftrightarrow{BC}$.

Theorem 4.6.6. *Every parallelogram is a convex quadrilateral.*

Proof. Exercise 2. □

Theorem 4.6.7. *If $\triangle ABC$ is a triangle, D is between A and B, and E is between A and C, then $\square BCED$ is a convex quadrilateral.*

Proof. Exercise 5. (See Figure 4.27.) □

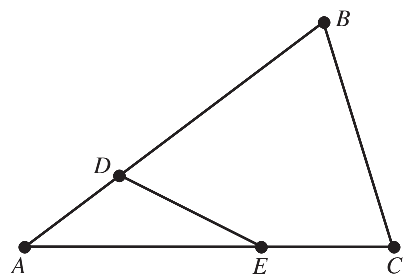

FIGURE 4.27: $\square DECB$ is a convex quadrilateral

Theorem 4.6.8. *The quadrilateral $\square ABCD$ is convex if and only if the diagonals $\overline{AC}$ and $\overline{BD}$ have an interior point in common.*

Proof. Assume, first, that $\square ABCD$ is a convex quadrilateral (hypothesis). We must prove that the diagonals $\overline{AC}$ and $\overline{BD}$ have a point in common. Observe that C is in the interior of $\angle DAB$ (definition of convex quadrilateral). Hence $\overrightarrow{AC}$ intersects $\overline{BD}$ in a point we will call E (Crossbar Theorem). It follows in a similar way that $\overrightarrow{BD}$ intersects $\overline{AC}$ in a point E'. But $\overleftrightarrow{AC}$ and $\overleftrightarrow{BD}$ meet in at most one point (Incidence Postulate) and so $E = E'$ and E lies on both $\overline{AC}$ and $\overline{BD}$.

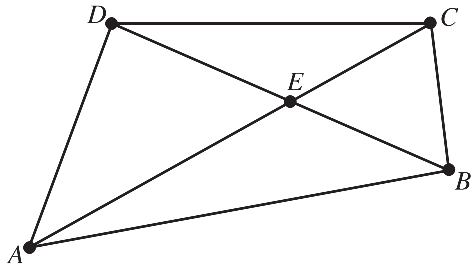

FIGURE 4.28: The diagonals $\overline{AC}$ and $\overline{BD}$ intersect at the point E

The proof of the converse is left as an exercise (Exercise 6). □

Corollary 4.6.9. *If $\square ABCD$ and $\square ACBD$ are both quadrilaterals, then $\square ABCD$ is not convex. If $\square ABCD$ is a nonconvex quadrilateral, then $\square ACBD$ is a quadrilateral.*

Proof. Exercise 7. □

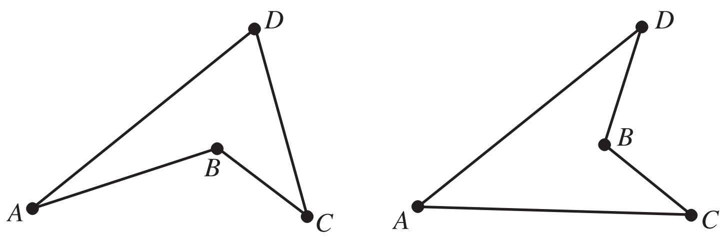

FIGURE 4.29: $\square ABCD$ and $\square ACBD$ are both quadrilaterals

EXERCISES 4.6

1. Prove that the angle sum of any convex quadrilateral is $\leqslant 360°$ (Theorem 4.6.4).
2. Prove that every parallelogram is convex (Theorem 4.6.6).
3. An alternate definition of convex quadrilateral. The opposite sides $\overline{AB}$ and $\overline{CD}$ of a quadrilateral $\square ABCD$ are said to be *semiparallel* if $\overline{AB} \cap \overleftrightarrow{CD} = \emptyset$ and $\overline{CD} \cap \overleftrightarrow{AB} = \emptyset$. Prove the following two results.
   (a) If one pair of opposite sides of $\square ABCD$ is semiparallel, then the other is.
   (b) $\square ABCD$ is convex if and only if both pairs of opposite sides are semiparallel.
4. A quadrilateral $\square ABCD$ is called a *trapezoid* if either $\overleftrightarrow{AB} \parallel \overleftrightarrow{CD}$ or $\overleftrightarrow{BC} \parallel \overleftrightarrow{AD}$. Prove that every trapezoid is convex.
5. Prove Theorem 4.6.7.
6. Prove that a quadrilateral is convex if the diagonals have a point in common (the remaining part of Theorem 4.6.8).

7. Prove Corollary 4.6.9.

8. A *rhombus* is a quadrilateral that has four congruent sides. Prove the following theorems about rhombi.
(a) Every rhombus is convex.
(b) The diagonals of a rhombus intersect at a point that is the midpoint of each diagonal.
(c) The diagonals of a rhombus are perpendicular.
(d) Every rhombus is a parallelogram.

9. Existence of rhombi. Squares do not necessarily exist in neutral geometry, but rhombi exist in profusion. Prove this by showing that if $\overline{AB}$ and $\overline{CD}$ are two segments that share a common midpoint and $\overleftrightarrow{AB} \perp \overleftrightarrow{CD}$, then $\square ACBD$ is a rhombus.

10. Let $\square ABCD$ be a convex quadrilateral. Prove that each of the following conditions implies that $\square ABCD$ is a parallelogram.
(a) $\triangle ABC \cong \triangle CDA$.
(b) $AB = CD$ and $BC = AD$.
(c) $\angle DAB \cong \angle BCD$ and $\angle ABC \cong \angle CDA$.
(d) The diagonals $\overline{AC}$ and $\overline{BD}$ share a common midpoint.

11. Kites and darts. Let $\square ABCD$ be a quadrilateral such that $AB = AD$ and $CB = CD$. If such a quadrilateral is convex, it is called a *kite*; if it is not convex, it is called a *dart*. Draw pictures of both cases to discover the reason for the names. Prove that $\angle ABC \cong \angle ADC$ in each case.

4.7 STATEMENTS EQUIVALENT TO THE EUCLIDEAN PARALLEL POSTULATE

As was pointed out earlier, Euclid delayed using his fifth postulate until he absolutely needed it. Our exploration of neutral geometry paralleled that of Euclid through the Alternate Interior Angles Theorem and its corollaries (Euclid's Propositions 27 and 28). After those theorems our treatment diverged from Euclid and we proved the Saccheri-Legendre Theorem regarding angle sums. In that theorem we avoided the use of any parallel postulate by proving a weaker theorem than Euclid did. We now take a different tack. We return to Euclid and consider his Proposition 29, which is the converse to the Alternate Interior Angles Theorem. We will see that Euclid was exactly right to invoke his Fifth Postulate at that point. We will prove, in fact, that the Fifth Postulate is logically equivalent to Proposition 29.

It is important to understand clearly what it means to say that the two statements are logically equivalent in this context. It means that either one of the statements could be assumed as an axiom and then the other could be proved as a theorem. More specifically, it means that if we start with all the axioms of neutral geometry and add one of the statements as an additional axiom, then the other statement can be proved as a theorem. This is often misunderstood. We are not claiming that these statements are logically equivalent in isolation. For example, Euclid's Fifth Postulate is a correct statement on the sphere, but the Euclidean Parallel Postulate is not. Hence the two statements cannot be logically equivalent by themselves. What we will show in this section is that they are logically equivalent in the presence of the axioms of neutral geometry. Thus adding one of the statements to our existing system of axioms is equivalent to adding the other.

Another way to understand this is in terms of models. When we assert that two statements are equivalent in neutral geometry we are saying that if one of the statements is true in a given model for neutral geometry, then the other must be true in that model as well.

We will prove that a surprisingly large number of the familiar theorems of Euclidean geometry are actually equivalent to the Parallel Postulate in the sense just explained. The approach taken in this section is consistent with what was done historically. The hope of many geometers was to prove Euclid's Fifth Postulate using only Euclid's other postulates. Failing that, the next best thing was to assume a simpler, more intuitively obvious postulate instead and then to demonstrate that Euclid's first four postulates together with the new postulate could be used to prove Euclid's Fifth Postulate. All such attempts to improve on Euclid have come to nothing in the sense that the new postulates invariably turn out to be equivalent to Euclid's postulate.

Here is a statement of Euclid's Proposition 29.

Converse to the Alternate Interior Angles Theorem. *If two parallel lines are cut by a transversal, then both pairs of alternate interior angles are congruent.*

The Converse to the Alternate Interior Angles Theorem is *not* a theorem in neutral geometry. But we do have the following theorem.

Theorem 4.7.1. *The Converse to the Alternate Interior Angles Theorem is equivalent to the Euclidean Parallel Postulate.*

Pay particular attention to the logical structure of the proof of this theorem. The proof itself is part of neutral geometry even though neither of the statements being proved is a theorem in neutral geometry. In the first half of the proof, the entire Converse to the Alternate Interior Angles Theorem is assumed as part of the hypothesis. The "givens" of the Euclidean Parallel Postulate (the line ℓ and the external point P) are also included as hypotheses. In the second half of the proof we prove the Converse to the Alternate Interior Angles Theorem. For that part of the proof, the hypotheses consist of the entire Euclidean Parallel Postulate along with the hypotheses of the Converse to the Alternate Interior Angles Theorem. In both parts of the proof an unstated hypothesis is the assumption that all the axioms of neutral geometry hold.

Proof of Theorem 4.7.1. First assume the Converse to the Alternate Interior Angles Theorem (hypothesis). Let ℓ be a line and let P be an external point (hypothesis). We must prove that there is exactly one line m such that P lies on m and $m \parallel \ell$. Construct a parallel line m and perpendicular transversal using the Double Perpendicular Construction (page 84). We must prove that m is the only such line. Suppose m' is any line such that P lies on m' and $m' \parallel \ell$. Then t is a transversal for ℓ and m' so the angle made by t and m' must be equal to the angle made by t and ℓ (Converse to the Alternate Interior Angles Theorem). Hence $m' \perp t$ and so $m' = m$ (the uniqueness part of Protractor Postulate, Part 3).

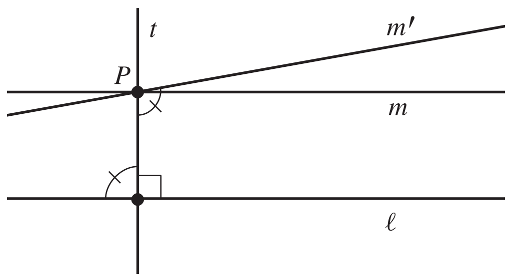

FIGURE 4.30: Converse to the Alternate Interior Angles Theorem implies the Euclidean Parallel Postulate

Now assume the Euclidean Parallel Postulate (hypothesis). Suppose ℓ and ℓ' are parallel lines with transversal t (hypothesis). We must prove that both pairs of alternate interior angles are congruent. Let B' denote the point at which t intersects ℓ'. Let ℓ'' be the line through B' for which the alternate interior angles formed by ℓ and ℓ'' with transversal t are congruent. (This line exists by the Protractor Postulate, Part 3.) By the Alternate Interior Angles Theorem, $\ell'' \parallel \ell$. The uniqueness part of the Euclidean Parallel Postulate implies that $\ell'' = \ell'$. Hence the alternate interior angles formed by ℓ and ℓ' with transversal t are congruent. □

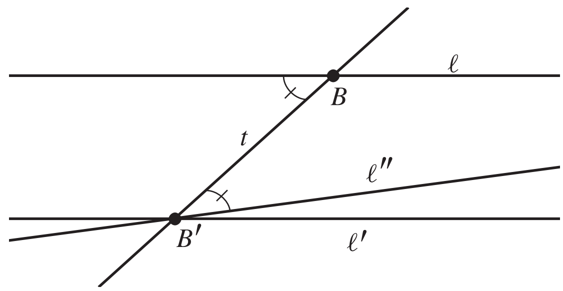

FIGURE 4.31: The Euclidean Parallel Postulate implies the converse to the Alternate Interior Angles Theorem

The axiom we have been calling the Euclidean Parallel Postulate is not exactly the same as Euclid's Fifth Postulate. The axiom we refer to as the Euclidean Parallel Postulate is often called *Playfair's Postulate* after the eighteenth-century Scottish mathematician John Playfair (1748–1819), even though it had been formulated much earlier by Proclus. Playfair stated the postulate in one of many attempts by various mathematicians to improve on Euclid by replacing Euclid's Fifth Postulate with something simpler. Of course Playfair's Postulate turned out to be logically equivalent to Euclid's. Let us state Euclid's Fifth Postulate in the language of this course.

Euclid's Postulate V. *If ℓ and ℓ' are two lines cut by a transversal t in such a way that the sum of the measures of the two interior angles on one side of t is less than* $180°$*, then ℓ and ℓ' intersect on that side of t.*

Theorem 4.7.2. *Euclid's Postulate V is equivalent to the Euclidean Parallel Postulate.*

Proof. First assume the Euclidean Parallel Postulate (hypothesis). Let ℓ and ℓ' be two lines cut by a transversal t such that the sum of the measures of the two interior angles on one side of t is less than $180°$ (hypothesis). Let B and B' be the points where t cuts ℓ and ℓ', respectively. There is a line ℓ'' such that B' lies on ℓ'' and both pairs of nonalternating interior angles formed by ℓ and ℓ'' with transversal t have measures whose sum is $180°$ (by the existence part of Protractor Postulate, Part 3). Note that $\ell'' \neq \ell'$ (by the uniqueness part of Protractor Postulate, Part 3) and $\ell'' \parallel \ell$ (Alternate Interior Angles Theorem). Hence ℓ' is not parallel to ℓ (by the uniqueness part of the Euclidean Parallel Postulate). Thus there exists a point C that lies on both ℓ and ℓ' (negation of the definition of parallel). Now $\triangle BB'C$ is a triangle and so the sum of the measures of any two interior angles must be less than $180°$ (Lemma 4.5.3). It follows that C must be on the side of t where the interior angles formed by ℓ and ℓ' with transversal t have measures whose sum is less than $180°$.

The other half of the proof is left as an exercise (Exercise 1). □

Hilbert also has a parallel postulate. It is the same as the Euclidean Parallel Postulate except that he only asserts that there is at most one parallel line, rather than exactly one. Since we have already proved that there is at least one parallel line, it is obvious that Hilbert's Parallel Postulate is equivalent to the Euclidean Parallel Postulate. Hilbert

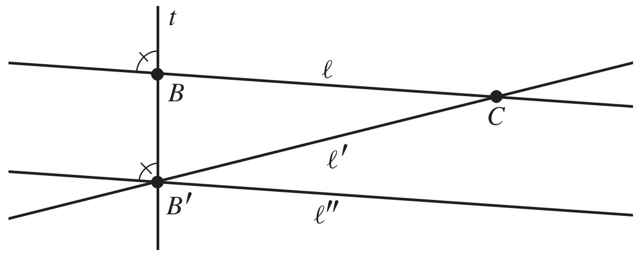

FIGURE 4.32: The Euclidean Parallel Postulate implies Euclid's Postulate V

was interested in finding a set of postulates that assumed no more than was absolutely necessary. We prefer to use Playfair's Postulate because it makes a clearer contrast with the Hyperbolic and Elliptic Parallel Postulates.

Hilbert's Parallel Postulate. *For every line ℓ and for every external point P there exists at most one line m such that P lies on m and $m \parallel \ell$.*

The next theorem gives several other statements that are equivalent to the Euclidean Parallel Postulate. Be sure to notice the fourth part of the theorem. It is tempting to assume that parallelism is transitive just on the basis of general principles, but that is not the case. Transitivity of Parallelism is Euclid's Proposition 30.

Theorem 4.7.3. *Each of the following statements is equivalent to the Euclidean Parallel Postulate.*

1. (Proclus's Axiom) *If ℓ and ℓ' are parallel lines and $t \neq \ell$ is a line such that t intersects ℓ, then t also intersects ℓ'.*
2. *If ℓ and ℓ' are parallel lines and t is a transversal such that $t \perp \ell$, then $t \perp \ell'$.*
3. *If ℓ, m, n and k are lines such that $k \parallel \ell$, $m \perp k$, and $n \perp \ell$, then either $m = n$ or $m \parallel n$.*
4. (Transitivity of Parallelism) *If ℓ is parallel to m and m is parallel to n, then either $\ell = n$ or $\ell \parallel n$.*

Proof. Exercises 2, 3, 4, and 5. □

There is a close relationship between the parallel postulate and angle sums. We will explore that relationship further in the next section, but for now we prove the basic equivalence. First let us state the usual assumption about angle sums as a postulate.

Angle Sum Postulate. *If $\triangle ABC$ is a triangle, then $\sigma(\triangle ABC) = 180°$.*

Theorem 4.7.4. *The Euclidean Parallel Postulate is equivalent to the Angle Sum Postulate.*

The proof relies on the following lemma. This lemma will also be important in Chapter 6 when we study parallelism in hyperbolic geometry.

Lemma 4.7.5. *Suppose $\overline{PQ}$ is a segment and Q' is a point such that $\angle PQQ'$ is a right angle. For every $\epsilon > 0$ there exists a point T on $\overrightarrow{QQ'}$ such that $\mu(\angle PTQ) < \epsilon°$.*

Proof. Let $\overline{PQ}$ be a segment and let Q' be a point such that $\angle PQQ'$ is a right angle (hypothesis). Let $\epsilon > 0$ be given (hypothesis). Choose a point P', on the same side of $\overleftrightarrow{PQ}$ as Q', such that $\overleftrightarrow{PP'} \perp \overleftrightarrow{PQ}$. Observe that every point T on $\overrightarrow{QQ'}$ is in the interior of $\angle QPP'$ (definition of interior). Choose a sequence of points $T_1, T_2, \ldots$ on $\overrightarrow{QQ'}$ as follows.

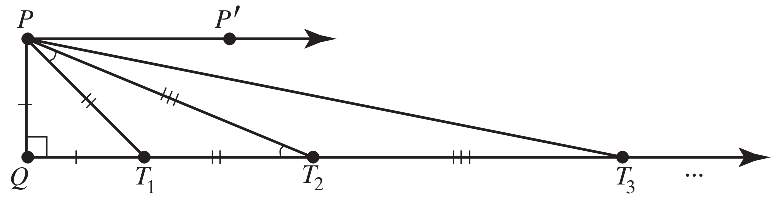

FIGURE 4.33: The construction of $T_1, T_2, \ldots$

First choose T_1 such that $\overline{PQ} \cong \overline{QT_1}$. Then choose T_2 such that $\overline{PT_1} \cong \overline{T_1T_2}$. Inductively choose T_n such that T_{n-1} is between Q and T_n and such that $\overline{PT_{n-1}} \cong \overline{T_{n-1}T_n}$. Points with those properties can be chosen using the Ruler Postulate.

Now $\angle QT_nP \cong \angle T_{n-1}PT_n$ for every n (Isosceles Triangle Theorem). For each n, Part 4 of the Protractor Postulate gives

$$\mu(\angle QPT_1) + \mu(\angle T_1PT_2) + \cdots + \mu(\angle T_{n-1}PT_n)$$
$$= \mu(\angle QPT_n) < \mu(\angle QPP') = 90^\circ.$$

Suppose $\mu(\angle T_{i-1}PT_i) \geqslant \epsilon$ for every i (RAA hypothesis). By the Archimedean Property of Real Numbers (Axiom E.3.4) there is an n for which $n\epsilon > 90$. It follows that

$$\mu(\angle QPT_1) + \mu(\angle T_1PT_2) + \cdots + \mu(\angle T_{n-1}PT_n) \geqslant n\epsilon^\circ > 90^\circ.$$

This contradicts the earlier computation. Hence we must reject the RAA hypothesis and can conclude that there exists an i such that $\mu(\angle T_{i-1}PT_i) < \epsilon^\circ$. Therefore, $\mu(\angle QT_iP) < \epsilon^\circ$ and so $T = T_i$ satisfies the conclusion of the lemma. □

Another way to formulate the last proof is to use two facts from calculus in place of the Archimedean Property. Specifically, we could use the fact that if an infinite series has positive terms and is bounded, then it converges, together with the fact that the nth term in a convergent series must approach zero.

Proof of Theorem 4.7.4 . The proof that the Euclidean Parallel Postulate implies that the angle sum of every triangle $\triangle ABC$ equals 180° is left as an exercise (Exercise 6).

For the converse, we must prove that if the angle sum for every triangle is 180°, then the Euclidean Parallel Postulate holds. We will actually prove the contrapositive; we will prove that if there exists a line ℓ and an external point P through which there are multiple parallel lines, then there is a triangle whose angle sum is different from 180°.

Let ℓ be a line and let P be an external point such that there is more than one line through P that is parallel to ℓ (hypothesis). Drop a perpendicular from P to ℓ and call the foot of that perpendicular Q. As usual, let m be the line through P that is perpendicular to $\overleftrightarrow{PQ}$. Then $m \parallel \ell$. By our hypothesis on ℓ and P, there exists another line m', different from m, such that P lies on m' and $m' \parallel \ell$. Choose a point S on m' such that S is on the same side of m as Q. Choose a point R on m such that R is on the same side of $\overleftrightarrow{PQ}$ as S. Finally, choose a point T on ℓ such that T lies on the same side of $\overleftrightarrow{PQ}$ as S and $\mu(\angle QTP) < \mu(\angle SPR)$ (Lemma 4.7.5). We will prove that $\sigma(\triangle PQT) < 180^\circ$.

Now T is in the interior of $\angle QPS$ (otherwise $\overrightarrow{PS}$ would have to meet $\overline{QT}$ by the Crossbar Theorem). Therefore, $\mu(\angle QPT) < \mu(\angle QPS)$ (Protractor Postulate). The

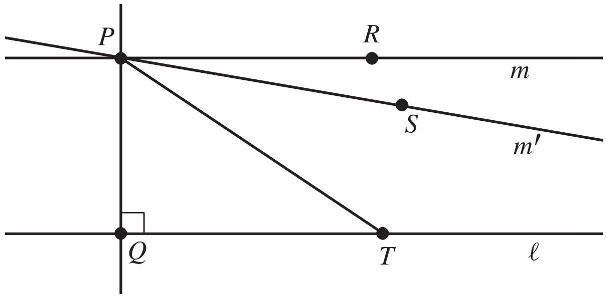

FIGURE 4.34: $\triangle PQT$ has angle sum less than 180°

way in which R and S were chosen guarantees that S is in the interior of $\angle QPR$, and so $\mu(\angle SPR) + \mu(\angle SPQ) = \mu(\angle RPQ)$ (Protractor Postulate). Hence

$$\begin{aligned}\sigma(\triangle QTP) &= \mu(\angle PQT) + \mu(\angle QTP) + \mu(\angle TPQ) \\ &< \mu(\angle PQT) + \mu(\angle SPR) + \mu(\angle QPS) \\ &= \mu(\angle PQT) + \mu(\angle RPQ) \\ &= 180^\circ\end{aligned}$$

because angles $\angle PQT$ and $\angle RPQ$ are right angles. □

We end this section with an equivalence between the existence of (noncongruent) similar triangles and the Euclidean Parallel Postulate. A mathematician named John Wallis (1616–1703) took the existence of similar triangles as his postulate and used it to prove Euclid's Fifth Postulate. Wallis's Postulate also proved to be equivalent to Euclid's Fifth Postulate.

Definition 4.7.6. Triangles $\triangle ABC$ and $\triangle DEF$ are *similar* if $\angle ABC \cong \angle DEF$, $\angle BCA \cong \angle EFD$, and $\angle CAB \cong \angle FDE$.

Notation. Write $\triangle ABC \sim \triangle DEF$ if $\triangle ABC$ is similar to $\triangle DEF$.

Wallis's Postulate. *If $\triangle ABC$ is a triangle and $\overline{DE}$ is a segment, then there exists a point F such that $\triangle ABC \sim \triangle DEF$.*

Theorem 4.7.7. *Wallis's Postulate is equivalent to the Euclidean Parallel Postulate.*

Proof. First assume the Euclidean Parallel Postulate (hypothesis). Let $\triangle ABC$ be a triangle and let $\overline{DE}$ be a segment (hypothesis). There exists a ray $\overrightarrow{DG}$ such that $\angle EDG \cong \angle BAC$ and there exists a ray $\overrightarrow{EH}$, with H on the same side of $\overleftrightarrow{DE}$ as G, such that $\angle DEH \cong \angle ABC$ (Protractor Postulate, Part 3). There exists a point F where $\overrightarrow{DG}$ and $\overrightarrow{EH}$ meet (Euclid's Postulate V). It follows from the Euclidean Parallel Postulate and Theorem 4.7.4 that $\sigma(\triangle ABC) = \sigma(\triangle DEF) = 180^\circ$. Hence subtraction gives $\mu(\angle EFD) = \mu(\angle BCA)$. Therefore, $\triangle ABC \sim \triangle DEF$ (definition of similar).

Now assume Wallis's Postulate (hypothesis). Let ℓ be a line and let P be an external point (hypothesis). We must show that there exists exactly one line m such that P lies on m and $m \parallel \ell$. Drop a perpendicular from P to ℓ and call the foot of that perpendicular Q. Let m be the line through P that is perpendicular to $\overleftrightarrow{PQ}$. Let m' be a line such that P lies on m' and $m' \parallel \ell$. Suppose $m' \neq m$ (RAA hypothesis). Choose a point S on m' such

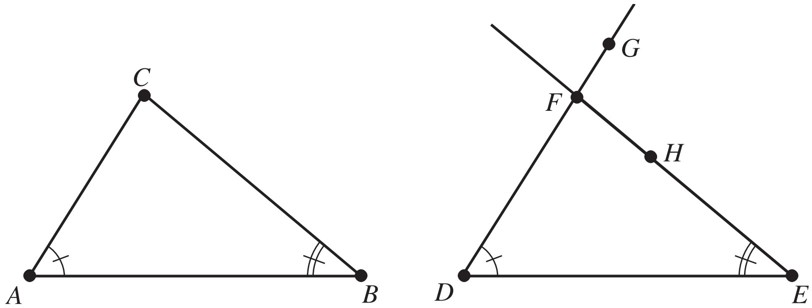

FIGURE 4.35: The Euclidean Parallel Postulate implies Wallis's Postulate

that S and Q are on the same side of m. Drop a perpendicular from S to $\overleftrightarrow{PQ}$ and call the foot of that perpendicular R. By Wallis's Postulate there exists a point T such that $\triangle PRS \sim \triangle PQT$. Since $\angle PQT$ is a right angle, T must lie on ℓ (Protractor Postulate, Part 3). Since $\angle QPT \cong \angle RPS$, T must lie on m' (Protractor Postulate, Part 3). Hence T is a point of $m' \cap \ell$ and so $m' \nparallel \ell$ (definition of parallel). This last statement contradicts the choice of m' so we must reject the RAA hypothesis and conclude that the parallel m is unique. □

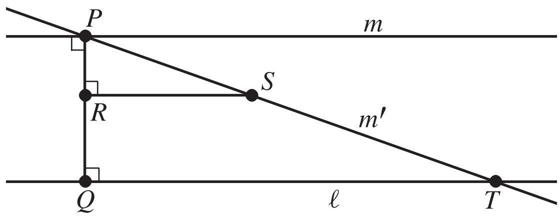

FIGURE 4.36: Wallis's Postulate implies the Euclidean Parallel Postulate

EXERCISES 4.7

1. Prove that Euclid's Postulate V implies the Euclidean Parallel Postulate (the second half of Theorem 4.7.2).
2. Prove that the Euclidean Parallel Postulate is equivalent to the first statement in Theorem 4.7.3.
3. Prove that the Euclidean Parallel Postulate is equivalent to the second statement in Theorem 4.7.3.
4. Prove that the Euclidean Parallel Postulate is equivalent to the third statement in Theorem 4.7.3.
5. Prove that the Euclidean Parallel Postulate is equivalent to the fourth statement in Theorem 4.7.3.
6. Prove that the Euclidean Parallel Postulate implies that the angle sum of any triangle is 180° (the second half of Theorem 4.7.4).

4.8 RECTANGLES AND DEFECT

In this section we explore angle sums for triangles in more depth. We have already seen that the Euclidean Parallel Postulate is equivalent to the assertion that every triangle has angle sum 180°. The main result of this section says that it is all or nothing: If just one triangle has angle sum equal to 180°, then every triangle has angle sum equal to 180°. Along the way toward the proof of that assertion we will be led to consider the question of whether or not a rectangle exists.

The Saccheri-Legendre Theorem asserts that every triangle has angle sum less than or equal to 180°. It allows for the possibility that an angle sum might fall short of the expected amount and we now want to seriously consider that possibility. It is convenient to give a name to the amount by which the angle sum might fall short.

Definition 4.8.1. For any triangle $\triangle ABC$, the *defect* of $\triangle ABC$ is defined by

$$\delta(\triangle ABC) = 180^\circ - \sigma(\triangle ABC).$$

By Saccheri-Legendre Theorem, the defect of every triangle is nonnegative. We will call a triangle *defective* if its defect is positive. Similarly, if $\square ABCD$ is a convex quadrilateral, define its defect by

$$\delta(\square ABCD) = 360^\circ - \sigma(\square ABCD).$$

The next theorem follows easily from Lemma 4.5.4. The simple additive relationship between the defect of a triangle and the defects of the triangles in a subdivision is the main reason that it is easier to work with defects than with angle sums.

Theorem 4.8.2 (Additivity of Defect).

1. *If $\triangle ABC$ is a triangle and E is a point in the interior of $\overline{BC}$, then*

$$\delta(\triangle ABC) = \delta(\triangle ABE) + \delta(\triangle ECA).$$

2. *If $\square ABCD$ is a convex quadrilateral, then*

$$\delta(\square ABCD) = \delta(\triangle ABC) + \delta(\triangle ACD).$$

Proof. Exercises 1 and 2. □

The following theorem is the main result of the section. The corollary that is stated immediately after the theorem is the result in which we are really interested, but it is convenient to state the theorem this way because it breaks the proof of the corollary down into manageable steps. The theorem as stated also clarifies the relationship between defect and the existence of rectangles.

Definition 4.8.3. A *rectangle* is a quadrilateral each of whose angles is a right angle.

Notice that the angle sum of a rectangle is 360° and so its defect is 0°. It follows from the Alternate Interior Angles Theorem and Theorem 4.6.6 that every rectangle is a convex quadrilateral. The following theorem shows that we cannot take the existence of rectangles for granted.

Theorem 4.8.4. *The following statements are equivalent.*

1. *There exists a triangle whose defect is* 0°.
2. *There exists a right triangle whose defect is* 0°.
3. *There exists a rectangle.*
4. *There exist arbitrarily large rectangles.*
5. *The defect of every right triangle is* 0°.
6. *The defect of every triangle is* 0°.

Corollary 4.8.5. *In any model for neutral geometry, there exists one triangle whose defect is* 0° *if and only if every triangle in that model has defect* 0°.

The statement "there exist arbitrarily large rectangles" means that for any given positive number M there exists a rectangle each of whose sides has length greater than M.

Before we tackle the proof of Theorem 4.8.4, we need a preliminary lemma.

Lemma 4.8.6. *If* $\triangle ABC$ *is any triangle, then at least two of the interior angles in* $\triangle ABC$ *are acute. If the interior angles at vertices* A *and* B *are acute, then the foot of the perpendicular from* C *to* $\overleftrightarrow{AB}$ *is between* A *and* B.

Proof. It follows from the Saccheri-Legendre Theorem that at least two of the interior angles must be acute. Let us say that the interior angles at vertices A and B are acute. Drop a perpendicular from C to $\overleftrightarrow{AB}$ and call the foot D. We must prove that D is between A and B.

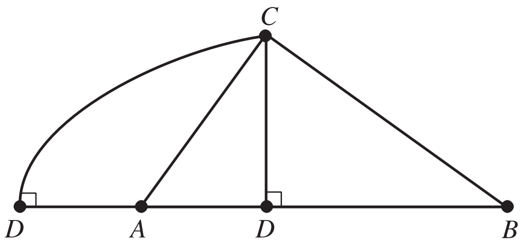

FIGURE 4.37: If D is not between A and B, the Exterior Angle Theorem is contradicted

First note that D cannot be equal to either A or B because if it were then the interior angle at that vertex would be a right angle. Suppose D is not between A and B (RAA hypothesis). Then either $D * A * B$ or $A * B * D$. If $D * A * B$, then the acute angle $\angle CAB$ is an exterior angle for $\triangle ACD$ and the right angle $\angle CDA$ is a remote interior angle for the same triangle. This contradicts the Exterior Angle Theorem. A similar proof shows that $A * B * D$ is impossible. Thus we can conclude that D is between A and B. □

Proof of Theorem 4.8.4. We will prove that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 6 \Rightarrow 1$. This cycle of implications shows that each of the statements implies all the others.

($1 \Rightarrow 2$) Let $\triangle ABC$ be a triangle such that $\delta(\triangle ABC) = 0^\circ$ (hypothesis). Relabel the vertices, if necessary, so that the interior angles at A and B are acute. By Lemma 4.8.6, there is a point D between A and B such that $\angle CDA$ is a right angle. Then triangles $\triangle ADC$ and $\triangle DBC$ are both right triangles. By Lemma 4.8.2, $\delta(\triangle ABC) = \delta(\triangle ADC) + \delta(\triangle DBC)$. Since all defects are nonnegative and $\delta(\triangle ABC) = 0^\circ$, it follows that $\delta(\triangle ADC) = \delta(\triangle DBC) = 0^\circ$.

($2 \Rightarrow 3$) Suppose $\triangle ABC$ is a triangle such that $\angle ABC$ is a right angle and $\delta(\triangle ABC) = 0^\circ$ (hypothesis). It follows that $\mu(\angle BAC) + \mu(\angle BCA) = 90^\circ$. There exists a point D, on the opposite side of $\overleftrightarrow{AC}$ from B, such that $\triangle CDA \cong \triangle ABC$ (Theorem 4.2.6). It is easy to check that $\square ABCD$ is a rectangle—see Figure 4.39.

($3 \Rightarrow 4$) The idea is to place two copies of the given rectangle next to each other to form a larger rectangle. This process is repeated to produce a rectangle whose dimensions are larger than any specified number—see Fig. 4.40.

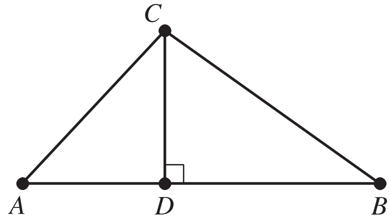

FIGURE 4.38: ($1 \Rightarrow 2$) Split the given triangle into two right triangles

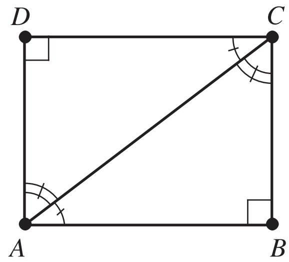

FIGURE 4.39: ($2 \Rightarrow 3$) Two copies of $\triangle ABC$ make a rectangle

Let $\square ABCD$ be a rectangle. Choose a point E on $\overrightarrow{AB}$ such that $A * B * E$ and $AB = BE$; then choose a point F on $\overrightarrow{DC}$ such that $D * C * F$ and $DC = CF$. By SAS, $\triangle ABC \cong \triangle EBC$. In particular, $\mu(\angle BCE) = \mu(\angle BCA) < 90°$, so E is in the interior of $\angle BCF$ (Theorem 3.4.5). The Angle Addition Postulate and subtraction give $\angle ACD \cong \angle ECF$, so a second application of SAS shows that $\triangle ADC \cong \triangle EFC$.

Now $\delta(\triangle ABC) + \delta(\triangle ACD) = \delta(\square ABCD)$ (Theorem 4.8.2), so $\delta(\triangle ABC) = \delta(\triangle ACD) = 0°$. (The defect of any rectangle is zero.) The fact that $\delta(\triangle ABC) = 0°$ means that $\mu(\angle BAC) + \mu(\angle BCA) = 90°$. Since $\square ABCD$ is a parallelogram, it is convex (Theorem 4.6.6). Hence C is in the interior of $\angle BAD$ and therefore $\mu(\angle BAC) + \mu(\angle CAD) = 90°$. Combining the last two statements gives $\angle DAC \cong \angle ACB$. A similar argument shows $\angle ACD \cong \angle CAB$. Therefore $\triangle ABC \cong \triangle CDA$ (ASA).

We now have $\triangle ABC \cong \triangle CDA \cong \triangle CFE$, so $\angle CFE$ is a right angle. A similar argument shows that $\angle BEF$ is a right angle. Therefore, $\square AEFD$ is a rectangle. One pair of sides of $\square AEFD$ is twice as long as the corresponding side of $\square ABCD$; specifically, $AE = 2AB = DF$. As suggested by Figure 4.40, we can apply the construction again to produce a rectangle $\square AEGH$ such that $AH = 2AD = EG$. By the Archimedean Property of Real Numbers, iteration of this process will produce arbitrarily large rectangles.

($4 \Rightarrow 5$) Let $\triangle ABC$ be a right triangle with right angle at vertex C (hypothesis). We must prove that $\delta(\triangle ABC) = 0°$.

By Statement 4, there exists a rectangle $\square DEFG$ such that $DG > AC$ and $FG > BC$. Choose a point B' on $\overline{GF}$ such that $GB' = CB$ and a point A' on $\overline{GD}$ such that $GA' = CA$. By SAS, $\triangle ABC \cong \triangle A'B'G$. In order to simplify notation, we will assume that $G = C$, $A' = A$, and $B' = B$.

Since $\square CDEF$ is a rectangle, it is convex (Theorem 4.6.6) and has defect $0°$. Hence $\delta(\triangle DEF) = \delta(\triangle CDF) = 0°$ (Additivity of defect, Theorem 4.8.2). Subdividing again gives $0° = \delta(\triangle CDF) = \delta(\triangle ADF) + \delta(\triangle AFC)$ (Theorem 4.8.2) and so $\delta(\triangle AFC) = 0°$. Subdividing once more gives $0° = \delta(\triangle AFC) = \delta(\triangle AFB) + \delta(\triangle ABC)$ and so

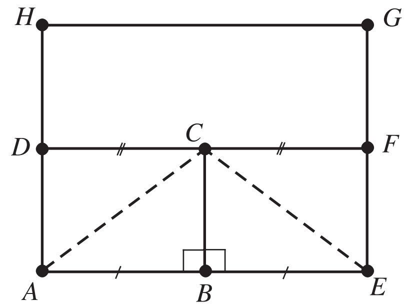

FIGURE 4.40: ($3 \Rightarrow 4$) Place copies of the rectangle next to each other to form a larger rectangle

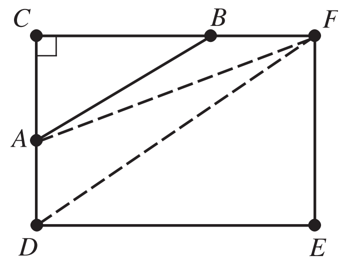

FIGURE 4.41: ($4 \Rightarrow 5$) Embed $\triangle ABC$ in a rectangle and use additivity of defect

$\delta(\triangle ABC) = 0°$.

($5 \Rightarrow 6$) By Lemma 4.8.6 the triangle can be subdivided into two right triangles. If each of the right triangles has defect $0°$, then Additivity of Defect implies that the original has defect $0°$ as well.

($6 \Rightarrow 1$) This implication is obvious. □

The last theorem indicates that the question of whether or not there might exist defective triangles can be answered by determining whether or not there exists a rectangle. As a result, many geometers have thought about the problem of the existence of rectangles. There have historically been many different approaches to the problem. The straightforward approach is simply to assume that rectangles exist. (This does, after all, seem intuitively obvious to most people.) That is the approach followed by Clairaut.

Clairaut's Axiom. *There exists a rectangle.*

By Theorem 4.8.4, Clairaut's Axiom implies that every triangle has defect $0°$ and thus implies the Euclidean Parallel Postulate (Theorem 4.7.4). On the other hand, the Euclidean Parallel Postulate implies that every triangle has angle sum $180°$, so the Euclidean Parallel Postulate implies Clairaut's Axiom. Hence we see that Clairaut's Axiom is also equivalent to the Euclidean Parallel Postulate.

Corollary 4.8.7. *Clairaut's Axiom is equivalent to the Euclidean Parallel Postulate.*

Many other mathematicians worked to construct rectangles. There are two standard ways to begin the construction of a rectangle. The quadrilaterals in the constructions have been named after two mathematicians who worked on the problem. Notice that both kinds of quadrilaterals described below "should" be rectangles in the sense that it would be easy to prove that they are rectangles using familiar theorems of Euclidean geometry.

Definition 4.8.8. A *Saccheri quadrilateral* is a quadrilateral $\square ABCD$ such that $\angle ABC$ and $\angle DAB$ are right angles and $\overline{AD} \cong \overline{BC}$. The segment $\overline{AB}$ is called the *base* of the Saccheri quadrilateral and the segment $\overline{CD}$ is called the *summit*. The two right angles $\angle ABC$ and $\angle DAB$ are called the *base angles*of the Saccheri quadrilateral and the angles $\angle CDA$ and $\angle BCD$ are called the *summit angles* of the Saccheri quadrilateral.

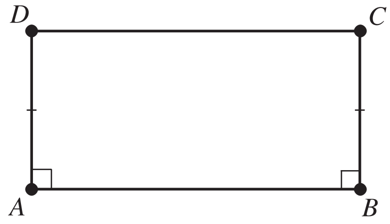

FIGURE 4.42: A Saccheri quadrilateral

It is easy to construct a Saccheri quadrilateral. Start with a line segment $\overline{AB}$ and erect perpendiculars at the endpoints. Choose points C and D, on the same side of $\overleftrightarrow{AB}$, such that $\overleftrightarrow{AD} \perp \overleftrightarrow{AB}$, $\overleftrightarrow{BC} \perp \overleftrightarrow{AB}$, and $\overline{AD} \cong \overline{BC}$.

Definition 4.8.9. A *Lambert quadrilateral* is a quadrilateral in which three of the angles are right angles.

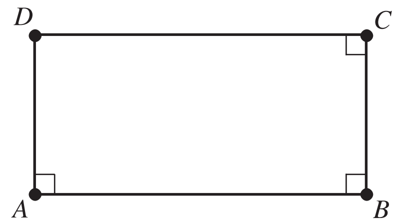

FIGURE 4.43: A Lambert quadrilateral

Again it is easy to construct a Lambert quadrilateral. Simply start with a right angle with vertex at B and choose a point D in the interior of the right angle. Drop perpendiculars from D to the two sides of the angle and call the feet of the perpendiculars A and C.

We have followed common practice in naming the two special types of quadrilaterals after Giovanni Saccheri (1667–1733) and Johann Lambert (1728–1777), respectively. This assignment of names appears, however, to reflect a western bias that does not adequately recognize the contributions of Islamic mathematicians. The Islamic world contributed much to geometry between the time of Euclid and the western Renaissance. In their exploration of Euclid's Fifth Postulate, Islamic mathematicians studied the very quadrilaterals we now name in honor of Saccheri. It has been contended that it would be more fitting to name them after the Persian geometer and poet Omar Khayyam (Umar al-Khayyami, 1048–1131) than Saccheri.[2]

The next two theorems spell out the important properties of Saccheri and Lambert quadrilaterals (at least those that can be proved in neutral geometry).

[2] See [23], for example.

Theorem 4.8.10 (Properties of Saccheri quadrilaterals). *If* $\square ABCD$ *is a Saccheri quadrilateral with base* $\overline{AB}$*, then*

1. *the diagonals* $\overline{AC}$ *and* $\overline{BD}$ *are congruent,*
2. *the summit angles* $\angle BCD$ *and* $\angle ADC$ *are congruent,*
3. *the segment joining the midpoint of* $\overline{AB}$ *to the midpoint of* $\overline{CD}$ *is perpendicular to both* $\overline{AB}$ *and* $\overline{CD}$*,*
4. $\square ABCD$ *is a parallelogram,*
5. $\square ABCD$ *is a convex quadrilateral, and*
6. *the summit angles* $\angle BCD$ *and* $\angle ADC$ *are either right or acute.*

Proof. Exercise 5. □

Theorem 4.8.11 (Properties of Lambert quadrilaterals). *If* $\square ABCD$ *is a Lambert quadrilateral with right angles at vertices A, B, and C, then*

1. $\square ABCD$ *is a parallelogram,*
2. $\square ABCD$ *is a convex quadrilateral,*
3. $\angle ADC$ *is either right or acute, and*
4. $BC \leqslant AD$.

Proof. Exercise 8. □

Both Saccheri and Lambert hoped to prove (in neutral geometry) that the quadrilaterals that bear their names are rectangles. Each of them succeeded only in ruling out the possibility that one of the angles is obtuse. Neither was able to eliminate the possibility that the unknown angles in their quadrilaterals might be acute. Thus their efforts to prove the Euclidean Parallel Postulate in this way failed. Despite that, their exploration of the properties of these quadrilaterals played an important part in the developments leading to the discovery of non-Euclidean geometry.

Saccheri and Lambert quadrilaterals continue to play a role in the study of parallelism. We will make extensive use of both kinds of quadrilaterals and their properties in later chapters when we investigate parallelism in more depth. Theorems 4.8.10 and 4.8.11 will be applied over and over again. Two sample applications are included in Exercises 6 and 7. The proof of the next theorem also illustrates what useful tools these quadrilaterals are. The result is known as Aristotle's Theorem because Proclus ascribed its statement to the philosopher Aristotle (see Heath [22], page 207).

Theorem 4.8.12 (Aristotle's Theorem). *If A, B, and C are three noncollinear points such that* $\angle BAC$ *is an acute angle and P and Q are two points on* $\overrightarrow{AB}$ *with* $A * P * Q$*, then* $d(P, \overleftrightarrow{AC}) < d(Q, \overleftrightarrow{AC})$*. Furthermore, for every positive number* d_0 *there exists a point R on* $\overrightarrow{AB}$ *such that* $d(R, \overleftrightarrow{AC}) > d_0$*.*

Proof. Exercise 10. □

EXERCISES 4.8

1. Prove Additivity of Defect for triangles (Part 1 of Theorem 4.8.2).
2. Prove Additivity of Defect for convex quadrilaterals (Part 2 of Theorem 4.8.2).
3. Prove the following converse to Lemma 4.8.6: *If* $\triangle ABC$ *is a triangle and the foot of the perpendicular from C to* $\overleftrightarrow{AB}$ *lies between A and B, then angles* $\angle CAB$ *and* $\angle CBA$ *are both acute.*

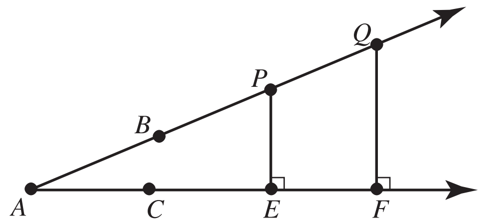

FIGURE 4.44: Aristotle's Theorem: $d(P, \ell)$ is an increasing function of AP and $d(P, \ell) \to \infty$ as $AP \to \infty$

4. Prove the following result. *If $\overline{AB}$ is the longest side of $\triangle ABC$, then the foot of the perpendicular from C to $\overleftrightarrow{AB}$ lies in the segment $\overline{AB}$.* Must the foot be between A and B?
5. Prove that every Saccheri quadrilateral has the properties listed in Theorem 4.8.10.
6. Prove the following result: *If ℓ and m are two distinct lines and there exist points P and Q on m such that $d(P, \ell) = d(Q, \ell)$, then either m and ℓ intersect at the midpoint of $\overline{PQ}$ or $m \parallel \ell$.*
7. Prove the following result: *If ℓ and m are two lines and there exist three distinct points P, Q, and R on m such that $d(P, \ell) = d(Q, \ell) = d(R, \ell)$, then either $m = \ell$ or $m \parallel \ell$.*
8. Prove that every Lambert quadrilateral has the properties listed in Theorem 4.8.11.
9. Let $\angle BAC$ be an acute angle and let P and Q be two points on $\overrightarrow{AB}$ such that $A * P * Q$ and $AP = PQ$. Prove that $d(Q, \overleftrightarrow{AC}) \geqslant 2d(P, \overleftrightarrow{AC})$.
10. Prove Aristotle's Theorem (Theorem 4.8.12).

4.9 THE UNIVERSAL HYPERBOLIC THEOREM

The Euclidean Parallel Postulate asserts that for every line ℓ and for every external point P there exists a unique parallel line through P, while the Hyperbolic Parallel Postulate asserts that for every line ℓ and for every external point P there are multiple parallel lines through P. Once one begins to think about this, an obvious question arises: Is it possible that for some lines and some external points there is a unique parallel while for others there are multiple parallels? The answer is that there are no mixed possibilities, that either there is a unique parallel in every situation or there are multiple parallels in every situation. The main theorem in the section asserts that if a model contains just one line and one external point where uniqueness of parallels fails, then uniqueness of parallels fails for every line and for every external point in that model. It is another all-or-nothing theorem, like the theorem regarding the existence of defective triangles.

Theorem 4.9.1 (The Universal Hyperbolic Theorem). *If there exists one line ℓ_0, an external point P_0, and at least two lines that pass through P_0 and are parallel to ℓ_0, then for every line ℓ and for every external point P there exist at least two lines that pass through P and are parallel to ℓ.*

Proof. Assume there exists a line ℓ_0, an external point P_0, and at least two lines that pass through P_0 and are parallel to ℓ_0 (hypothesis). This hypothesis implies that the Euclidean Parallel Postulate fails. Hence no rectangle can exist (Corollary 4.8.7).

Let ℓ be a line and let P an external point (hypothesis). We must prove that there exist at least two lines through P that are both parallel to ℓ. We begin by constructing one such line in the usual way. Drop a perpendicular to ℓ and call the foot of that perpendicular Q. Let m be the line through P that is perpendicular to $\overleftrightarrow{PQ}$. Choose a point R on ℓ that is different from Q and let t be the line through R that is perpendicular to ℓ. Drop a perpendicular from P to t and call the foot of that perpendicular S.

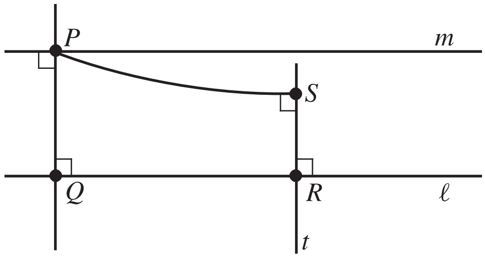

FIGURE 4.45: Proof of the Universal Hyperbolic Theorem

Now $\square PQRS$ is a Lambert quadrilateral. It cannot be a rectangle because, as noted in the first paragraph of this proof, the hypotheses of the theorem imply that no rectangle can exist. Hence $\angle QPS$ is not a right angle and $\overleftrightarrow{PS} \neq m$. But $\overleftrightarrow{PS}$ is parallel to ℓ (Alternate Interior Angles Theorem, applied using the transversal t). This completes the proof. $\square$

Notice that the hypothesis of the Universal Hyperbolic Theorem is just the negation of the Euclidean Parallel Postulate while the conclusion is the Hyperbolic Parallel Postulate. Therefore, the theorem can be restated in the following way.

Corollary 4.9.2. *The Hyperbolic Parallel Postulate is equivalent to the negation of the Euclidean Parallel Postulate.*

The following corollary is a fitting conclusion to the chapter.

Corollary 4.9.3. *In any model for neutral geometry either the Euclidean Parallel Postulate or the Hyperbolic Parallel Postulate will hold.*

In the next two chapters we will investigate the two possibilities separately.

SUGGESTED READING

Chapters 3 and 4 of *The Non-Euclidean Revolution*, [44].